

GenCore version 5.1.3

Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 3, 2002, 14:26:31 ; Search time 6.53333 Seconds  
(without alignments)  
34.123 Million cell updates/sec

Title: US-09-924-654-4\_COPY\_404\_417

Perfect score: 69

Sequence: 1 LILLRGLPGSGKTT 14

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 102317 seqs, 15924203 residues

Total number of hits satisfying chosen parameters: 102317

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	69	100.0	583	10	US-09-924-654-4
2	69	100.0	594	10	US-09-925-300-1079
3	50	72.5	92	10	US-09-795-693-29
4	50	72.5	92	10	US-09-795-693-30
5	50	72.5	92	10	US-09-795-693-42
6	50	72.5	191	10	US-09-815-242-11348
7	50	72.5	235	10	US-09-803-286A-4
8	50	72.5	479	10	US-09-971-309-64
9	49	71.0	448	10	US-09-815-242-11562
10	47	68.1	215	10	US-09-815-242-5542
11	47	68.1	224	10	US-09-815-242-12227
12	47	68.1	224	10	US-09-815-242-12824
13	47	68.1	457	10	US-09-815-242-11948
14	47	68.1	462	10	US-09-815-242-10965
15	46	66.7	65	10	US-09-756-998-6
16	46	66.7	453	10	US-09-815-242-10263
17	46	66.7	501	10	US-09-815-242-14089
18	46	66.7	748	10	US-09-949-192-43
19	46	66.7	751	10	US-09-815-242-5832

20	46	66.7	825	10	US-09-815-242-12963	Sequence 12963, A
21	46	66.7	869	10	US-09-815-242-10623	Sequence 10623, A
22	46	66.7	891	10	US-09-949-192-45	Sequence 45, Appl
23	46	66.7	897	10	US-09-949-192-49	Sequence 49, Appl
24	45	65.2	212	10	US-09-815-242-13219	Sequence 13219, A
25	45	65.2	213	10	US-09-815-242-4889	Sequence 4889, Ap
26	45	65.2	231	10	US-09-815-242-10536	Sequence 10536, A
27	45	65.2	472	10	US-09-815-242-4945	Sequence 4945, Ap
28	45	65.2	477	10	US-09-815-242-10861	Sequence 10861, A
29	44	63.8	191	9	US-09-895-913A-336	Sequence 336, App
30	44	63.8	213	10	US-09-815-242-10129	Sequence 10129, A
31	44	63.8	213	10	US-09-815-242-14070	Sequence 14070, A
32	44	63.8	215	10	US-09-815-242-5382	Sequence 5382, Ap
33	44	63.8	221	10	US-09-815-242-12195	Sequence 12195, A
34	44	63.8	307	9	US-10-108-605-183	Sequence 183, App
35	44	63.8	466	10	US-09-815-242-13525	Sequence 13525, A
36	44	63.8	523	10	US-09-815-242-13254	Sequence 13254, A
37	44	63.8	523	10	US-09-814-041A-2	Sequence 2, Appli
38	44	63.8	523	10	US-09-815-242-13695	Sequence 13695, A
39	44	63.8	726	10	US-09-815-242-10554	Sequence 10554, A
40	43	62.3	456	10	US-09-756-998-21	Sequence 21, Appl
41	43	62.3	456	10	US-09-840-787-5	Sequence 5, Appli
42	43	62.3	793	10	US-09-815-242-13689	Sequence 13689, A
43	43	62.3	806	10	US-09-922-217-1117	Sequence 1117, Ap
44	43	62.3	806	10	US-09-922-217-1118	Sequence 1118, Ap
45	43	62.3	1045	10	US-09-815-242-10617	Sequence 10617, A

ALIGNMENTS

RESULT 1  
US-09-924-654-4  
; Sequence 4, Application US/09924654  
; Patent No. US20020146712A1  
; GENERAL INFORMATION:  
; APPLICANT: Goold, Richard D.  
; APPLICANT: Akerblom, Ingrid E.  
; APPLICANT: Sellhamer, Jeffrey J.  
; APPLICANT: Coleman, Roger  
; TITLE OF INVENTION: TUMOR SUPPRESSOR  
; FILE REFERENCE: PC-0049 CIP  
; CURRENT APPLICATION NUMBER: US/09/924,654  
; CURRENT FILING DATE: 2001-08-07  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: PERL Program  
; SEQ ID NO 4  
; LENGTH: 583  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. US20020146712A1 496071CD1  
US-09-924-654-4

Query Match 100.0%; Score 69; DB 10; Length 583;  
Best Local Similarity 100.0%; Pred. No. 0.0033;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LILLRGLPGSGKTT 14

Db 404 LILLRGLPGSGKTT 417

RESULT 2

US-09-925-300-1079  
; Sequence 1079, Application US/09925300.  
; Patent No. US20020151681A1  
; GENERAL INFORMATION:  
; APPLICANT: Craig Rosen,  
; APPLICANT: Steve Ruben  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA101

```
; CURRENT APPLICATION NUMBER: US/09/925,300
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1079
; LENGTH: 594
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: SITE
; LOCATION: (430)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-300-1079

Query Match      100.0%; Score 69; DB 10; Length 594;
Best Local Similarity 100.0%; Pred. No. 0.0033;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LILLRGLPGSGKTT 14
Db 415 LILLRGLPGSGKTT 428

RESULT 3
US-09-795-693-29
; Sequence 29, Application US/09795693
; Patent No. US20020068710A1
; GENERAL INFORMATION:
; APPLICANT: Glucksmann, Maria A.
; TITLE OF INVENTION: 20685, 579, 17114, 23821, 33894, and
; FILE REFERENCE: 35800/209292
; CURRENT APPLICATION NUMBER: US/09/795,693
; CURRENT FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/185,906
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 92
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Pfam consensus sequence
US-09-795-693-29

Query Match      72.5%; Score 50; DB 10; Length 92;
Best Local Similarity 71.4%; Pred. No. 0.34;
Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 LILLRGLPGSGKTT 14
Db 4 VLLVGPSPGSGKTT 17

RESULT 4
US-09-795-693-30
; Sequence 30, Application US/09795693
; Patent No. US20020068710A1
; GENERAL INFORMATION:
; APPLICANT: Glucksmann, Maria A.
; TITLE OF INVENTION: 20685, 579, 17114, 23821, 33894, and
; FILE REFERENCE: 35800/209292
; CURRENT APPLICATION NUMBER: US/09/795,693
; CURRENT FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/185,906
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 42
```

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 92
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Pfam consensus sequence
US-09-795-693-30

Query Match      72.5%; Score 50; DB 10; Length 92;
Best Local Similarity 71.4%; Pred. No. 0.34;
Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 LILLRGLPGSGKTT 14
Db 4 VLLVGPSPGSGKTT 17

RESULT 5
US-09-795-693-42
; Sequence 42, Application US/09795693
; Patent No. US20020068710A1
; GENERAL INFORMATION:
; APPLICANT: Glucksmann, Maria A.
; TITLE OF INVENTION: 20685, 579, 17114, 23821, 33894, and
; FILE REFERENCE: 35800/209292
; CURRENT APPLICATION NUMBER: US/09/795,693
; CURRENT FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/185,906
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 42
; LENGTH: 92
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Pfam consensus sequence
US-09-795-693-42

Query Match      72.5%; Score 50; DB 10; Length 92;
Best Local Similarity 71.4%; Pred. No. 0.34;
Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 LILLRGLPGSGKTT 14
Db 4 VLLVGPSPGSGKTT 17

RESULT 6
US-09-815-242-11348
; Sequence 11348, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
```

```

; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11348
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-815-242-11348

Query Match      72.5%; Score 50; DB 10; Length 191;
Best Local Similarity 71.4%; Pred. No. 0.72; 3; Indels 0; Gaps 0;
Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 LILLRGLPGSGKTT 14
   | | | | | | | | | |
Db 4 LFLIIGAPGSGKTT 17
   | | | | | | | | | |

RESULT 7
US-09-803-286A-4
; Sequence 4, Application US/09803286A
; Patent No. US20020062504A1
; GENERAL INFORMATION:
; APPLICANT: Tanksley, Steven D.
; APPLICANT: Brommonschenkel, Sergio H.
; TITLE OF INVENTION: PLANT GENE CONFERRING RESISTANCE TO TOSPOVIRUSES
; FILE REFERENCE: 19603/3201
; CURRENT APPLICATION NUMBER: US/09/803,286A
; CURRENT FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: 60/188,356
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 235
; TYPE: PRT
; ORGANISM: Lycopersicon var.
US-09-803-286A-4

Query Match      72.5%; Score 50; DB 10; Length 235;
Best Local Similarity 64.3%; Pred. No. 0.89;
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 LILLRGLPGSGKTT 14
   | | | | | | | | | |
Db 4 VISITGMFPGSGKTT 17
   | | | | | | | | | |

RESULT 8
US-09-971-309-64
; Sequence 64, Application US/09971309
; Patent No. US20020106675A1
; GENERAL INFORMATION:
; APPLICANT: UEMORI, Takashi
; APPLICANT: SATO, Yoshiaki
; APPLICANT: FUJITA, Tomoko
; APPLICANT: MIYAKE, Kazue
; APPLICANT: MUKAI, Hiroyuki
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: DNA POLYMERASE-RELATED FACTORS
; FILE REFERENCE: 1422-0494P
; CURRENT APPLICATION NUMBER: US/09/971,309
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: US 09/446,504

```

```

; PRIOR APPLICATION NUMBER: 60/242, 578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253, 625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257, 931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269, 308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12227
; LENGTH: 224
; TYPE: PRF
; ORGANISM: Staphylococcus aureus
; US-09-815-242-12227

Query Match      68.1%; Score 47;
Best Local Similarity 69.2%; Pred. No. 2
Matches 9; Conservative 2; Mismatch

QY 2 ILRLGLPGSGKTT 14
|:|:|:|:|:|
Db 12 IILMLPGAGKGT 24

RESULT 12
US-09-815-242-12824
; Sequence 12824, Application US/09815242
; Patent No. US2002061569A1
; GENERAL INFORMATION:
; APPLICANT: Haseelbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Esses
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITFA 011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12824
; LENGTH: 224
; TYPE: PRF
; ORGANISM: Staphylococcus aureus
; US-09-815-242-12824

Query Match      68.1%; Score 47;
Best Local Similarity 69.2%; Pred. No. 2
Matches 9; Conservative 2; Mismatch

QY 2 ILRLGLPGSGKTT 14
|:|:|:|:|:|
Db 12 IILMLPGAGKGT 24

```

```
RESULT 13
US-09-815-242-11948
; Sequence 11948, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11948
; LENGTH: 457
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-11948

Query Match 68.1%; Score 47; DB 10; Length 457;
Best Local Similarity 64.3%; Pred. No. 4.9;
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 LLLRGLPGSGKTT 14
Db 102 VILMAGLQGAGKTT 115

RESULT 14
US-09-815-242-10965
; Sequence 10965, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
```

```
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10965
; LENGTH: 462
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-815-242-10965

Query Match 68.1%; Score 47; DB 10; Length 462;
Best Local Similarity 64.3%; Pred. No. 5;
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 LLLRGLPGSGKTT 14
Db 102 VILMAGLQGAGKTT 115

RESULT 15
US-09-756-998-6
; Sequence 6, Application US/09756998
; Patent No. US20010010931A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Steve
; APPLICANT: Lee, Jian Ming
; APPLICANT: Lightner, Jonathan
; APPLICANT: Odell, Joan
; TITLE OF INVENTION: PLANT GLYCOLYSIS AND RESPIRATION ENZYMES
; FILE REFERENCE: BB-1154
; CURRENT APPLICATION NUMBER: US/09/756,998
; CURRENT FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: 60/079,387
; PRIOR FILING DATE: 1998-03-16
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 6
; LENGTH: 65
; TYPE: PRT
; ORGANISM: Oryza sativa
US-09-756-998-6

Query Match 66.7%; Score 46; DB 10; Length 65;
Best Local Similarity 75.0%; Pred. No. 0.96;
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 LLRGLPGSGKTT 14
Db 35 LLHGPPGTGKTT 46

Search completed: December 3, 2002, 14:29:56
Job time : 6.53333 secs
```

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 3, 2002, 14:26:02 ; Search time 29.4 Seconds  
(without alignments)  
63.453 Million cell updates/sec

Title: US-09-924-654-4\_COPY\_404\_417  
Perfect score: 69  
Sequence: 1 LILLRGLPGSGKTT 14

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_101002.\*

```
1: /SID22/gcgdata/geneseq/geneseqp-embl/AA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseqp-embl/AA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseqp-embl/AA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseqp-embl/AA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseqp-embl/AA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseqp-embl/AA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseqp-embl/AA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseqp-embl/AA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseqp-embl/AA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseqp-embl/AA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseqp-embl/AA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseqp-embl/AA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseqp-embl/AA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseqp-embl/AA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseqp-embl/AA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseqp-embl/AA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseqp-embl/AA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseqp-embl/AA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseqp-embl/AA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseqp-embl/AA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseqp-embl/AA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseqp-embl/AA2001.DAT.*
23: /SID22/gcgdata/geneseq/geneseqp-embl/AA2002.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	69	100.0	119	23 ABP42992	Human ovarian anti
2	69	100.0	583	18 AAW26654	Human PANCIA assoc
3	69	100.0	594	21 AAB56501	Human prostate can
4	64	92.8	127	22 AAU28356	Novel human secret
5	64	92.8	243	22 AAB84326	Amino acid sequenc
6	59	85.5	659	22 AAB93057	Human protein sequ
7	56	81.2	394	23 ABB76271	Human IKK binding
8	56	81.2	629	22 AAY72168	Human RNA metaboli
9	55	79.7	241	22 AAU51368	Propionibacterium
10	51	73.9	314	21 AAB41858	Human ORFX ORF1622

11	51	73.9	592	22	ABG22415	Novel human diagno
12	50	72.5	20	21	AAV52063	P. horikoshi PHBNO
13	50	72.5	132	18	AAW55652	H. pylori ORF 06ep
14	50	72.5	191	18	AAW55730	H. pylori ORF 06ep
15	50	72.5	191	19	AAW98583	H. pylori GHPO 68
16	50	72.5	191	22	AAU35755	Helicobacter pylori
17	50	72.5	365	21	AAU90778	P. horikoshi PHBNO
18	50	72.5	468	21	AAV52008	P. horikoshi PHBNO
19	50	72.5	468	21	AAV51637	P. horikoshi PHB
20	50	72.5	479	20	AAW97108	Thermotable polyp
21	50	72.5	479	22	AAW96088	Putative P. abyssi
22	50	72.5	479	22	AAW62022	P. furiosus recomb
23	50	72.5	562	22	ABW68020	Drosophila melanog
24	50	72.5	722	22	AAU62951	Propionibacterium
25	50	72.5	795	22	ABW61567	Drosophila melanog
26	50	72.5	806	15	AAW55692	hnrNP U protein.
27	50	72.5	811	23	ABF41919	Human ovarian anti
28	50	72.5	1119	22	ABW58575	Drosophila melanog
29	50	72.5	1206	19	AAW47080	Tomato Mi resistan
30	50	72.5	1206	19	AAW53583	Wild tomato Mi res
31	50	72.5	1255	19	AAW55974	Tomato pest resist
32	50	72.5	1257	19	AAW55975	Tomato pest resist
33	50	72.5	1257	19	AAW53582	Wild tomato Mi res
34	50	72.5	1615	22	AAW62021	P. furiosus genom
35	49	71.0	365	18	AAW20643	H. pylori cytoplas
36	49	71.0	388	23	ABW91273	Herbicidally activ
37	49	71.0	448	22	AAU35969	Helicobacter pylori
38	48	69.6	172	22	AAW74678	Human protease and
39	48	69.6	178	22	AAW73594	Human colon cancer
40	48	69.6	190	22	AAW79854	Corynebacterium gl
41	48	69.6	448	20	AAV34715	Chlamydia pneumoni
42	48	69.6	526	21	AAW54362	Arabidopsis thalia
43	48	69.6	533	21	AAW54361	Arabidopsis thalia
44	48	69.6	649	21	AAW54273	Arabidopsis thalia
45	48	69.6	652	21	AAW54272	Arabidopsis thalia

#### ALIGNMENTS

RESULT 1  
ABP42992  
ID ABP42992 standard; Protein: 119 AA.  
XX ABP42992;  
AC ABP42992;  
XX  
DT 22-AUG-2002 (first entry)  
XX  
DE Human ovarian antigen HPDWO61, SEQ ID NO:4124.  
XX  
KW Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;  
KW ovarian cancer; breast cancer; tumour; reproductive system disorder;  
KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;  
KW PCOS; ovarian cyst; dysmenorrhea; endocrine disorder; infection;  
KW inflammatory condition; immune disorder; blood disorder;  
KW cardiovascular disorder; respiratory disorder; neurological disorder;  
KW gastrointestinal disorder; urinary system disorder; drug screening;  
KW gene therapy; chromosome mapping; forensic analysis;  
KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;  
KW antiinflammatory; gynaecological; reproductive.  
XX  
OS Homo sapiens.  
XX  
FN WO200200677-A1.  
XX  
PD 03-JAN-2002.  
XX  
PF 07-JUN-2001; 2001WO-US18569.  
XX  
PR 07-JUN-2000; 2000US-209467P.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX

PI Birse CE, Rosen CA;  
 XX WPI: 2002-147878/19.  
 DR N-PSDB; ABQ56069.  
 XX

Isolated nucleic acid molecules encoding novel ovarian polypeptides, useful in the prevention, treatment and diagnosis of cancer (e.g. ovarian cancer), immune disorders, cardiovascular disorders and neurological diseases -

XX Claim 11; SEQ ID No 4124; 2922pp; English.

XX The invention relates to 2175 novel human ovarian antigens (ABP41054-ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also encompasses polypeptides 90% identical and polynucleotides 95% identical to the sequences of the invention. The invention additionally relates to recombinant vectors and host cells comprising human ovarian antigen polynucleotides, antibodies against human ovarian antigens, and the use of ovarian antigen polynucleotides and polypeptides in diagnosing, treating, prognosing or preventing various ovary and/or breast-related disorders. Such conditions include ovarian cancer and breast cancer, and metastatic tumours of ovarian or breast origin, reproductive system disorders (e.g., infertility, disorders of pregnancy, anovulation, polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and vaginitis), immune disorders (e.g., congenital and acquired immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus), blood-related disorders (e.g., anaemia), cardiovascular disorders, respiratory disorders, neurological disorders, gastrointestinal disorders and urinary system disorders. Ovarian antigen polypeptides and polynucleotides may also be used in screening for compounds which modulate ovarian antigen expression or activity. The polynucleotides may further be used for gene therapy, chromosome mapping, in the identification of individuals and in forensic analysis, and the polypeptides may be used as food additives or to prepare antibodies useful in disease diagnosis, drug targeting and phenotyping. The present sequence represents a human ovarian antigen of the invention.

CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 119 AA;  
 Query Match 100.0%; Score 69; DB 23; Length 119;  
 Best Local Similarity 100.0%; Pred. No. 0.0019;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LILLRGLPGSGKTT 14  
 |||||  
 DB 57 LILLRGLPGSGKTT 70

RESULT 2  
 AAW26654  
 ID AAW26654 standard; Protein; 583 AA.  
 XX AAW26654;  
 AC AAW26654;  
 XX 16-FEB-1998 (first entry)  
 DT  
 XX Human PANC1A associated with pancreatic cancer.  
 DE PANC1A; pancreas cancer; diagnosis; therapy; human.  
 XX  
 KW Homo sapiens.  
 OS  
 XX WO9724437-A1.  
 PN  
 XX 10-JUL-1997.  
 PD  
 XX 18-DEC-1996; 96WO-US19981.  
 PF  
 XX

PR 15-MAR-1996; 96US-0616392.  
 PR 29-DEC-1995; 95US-0581240.  
 XX  
 PA (INCY-) INCYTE PHARM INC.  
 XX  
 PI Akerblom IE, Coleman R, Goold RD, Seilhamer JJ;  
 XX WPI: 1997-363673/33.  
 DR N-PSDB; AAT91010.  
 DR  
 XX DNA encoding PANC1A and PANC1B associated with pancreatic cancer - useful for diagnosis of pancreatic cancer and identifying risk of cancer  
 PT  
 PT Disclosure; Page 46-47; 65pp; English.  
 XX  
 PS This protein comprises human PANC1A, which is associated with pancreatic cancer and possibly other cancers. Its sequence was deduced from a full-length PANC1A cDNA clone (see AAT91010) derived from human teratocarcinoma hNT-2 cells. Specific fragments of PANC1A (see AAT91008) and PANC1B (see AAT91009) cDNA are claimed. Also claimed are: antisense molecules; expression vectors; transformed host cells; and purified PANC1A and PANC1B polypeptides. Diagnostic testing for the presence of PANC1A or PANC1B in a sample provides a means for positively correlating abnormal levels of PANC1A or PANC1B with pancreatic cancer (claimed). The host cells can be cultured for the recombinant production of PANC1A or PANC1B proteins (claimed). PANC1A and PANC1B polypeptides can be used to screen for antagonists and inhibitors, and to raise diagnostic antibodies.

XX SQ Sequence 583 AA;  
 Query Match 100.0%; Score 69; DB 18; Length 583;  
 Best Local Similarity 100.0%; Pred. No. 0.0092;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LILLRGLPGSGKTT 14  
 |||||  
 DB 404 LILLRGLPGSGKTT 417

RESULT 3  
 AAB56501  
 ID AAB56501 standard; Protein; 594 AA.  
 XX AAB56501;  
 AC AAB56501;  
 XX 13-MAR-2001 (first entry)  
 DT  
 XX Human prostate cancer antigen protein sequence SEQ ID NO:1079.  
 DE  
 XX Human; prostate cancer; prostate cancer antigen; detection; diagnosis; neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular; vulnary; gastrointestinal; nephrotropic; antiinfective; gynaecological; antibacterial; gene therapy; neural; immune; reproductive; renal; gastrointestinal; pulmonary; cardiovascular; proliferative disorder; wound; infectious disease.  
 KW  
 XX Homo sapiens.  
 OS  
 XX WO200055174-A1.  
 PN  
 XX 21-SEP-2000.  
 PD  
 XX 08-MAR-2000; 2000WO-US05988.  
 PF  
 XX 12-MAR-1999; 99US-0124270.  
 PR  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA (ROSE/) ROSEN C A.  
 PA  
 XX Rosen CA, Ruben SM;  
 PI

XX WPI; 2000-587513/55.  
 DR N-PSDB; AAF15704.  
 XX  
 XX Prostate cancer associated gene sequences, referred to as prostate  
 PT cancer antigens, useful for treatment, prevention, and diagnosis of  
 PT disorders such as prostate cancer -  
 PS  
 XX Claim 11; Page 1502-1504; 2338pp; English.  
 XX  
 CC AAF15566 to AAF16505 encode the human prostate cancer associated  
 CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.  
 CC The prostate cancer antigens can have neuroprotective, cytostatic,  
 CC cardioactive, immunomodulatory, muscular, vulnery, gastrointestinal,  
 CC nephrotropic, antiinfective, gynecological and antibacterial activities,  
 CC and can be used in gene therapy. The prostate cancer antigen  
 CC polynucleotides may be used for detection of prostate cancer, chromosome  
 CC identification, as chromosome markers, and for numerous other diagnostic  
 CC or research purposes. The prostate cancer antigens may be used to treat  
 CC disorders such as neural, immune, muscular, reproductive,  
 CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative  
 CC disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to  
 CC AAB57303 represent sequences used in the exemplification of the present  
 CC invention.  
 XX  
 XX Sequence 594 AA;  
 CC  
 CC Query Match 100.0%; Score 69; DB 21; Length 594;  
 CC Best Local Similarity 100.0%; Pred. No. 0.0094;  
 CC Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 Qy 1 LILLRGLPGSGKTT 14  
 Db |||||  
 415 LILLRGLPGSGKTT 428  
 RESULT 4  
 AAU28356  
 ID AAU28356 standard; Protein; 127 AA.  
 XX  
 AC AAU28356;  
 XX  
 XX 18-DEC-2001 (first entry)  
 XX  
 DE Novel human secretory protein, Seq ID No 713.  
 XX  
 KW Human; secreted protein; arthritis; Crohn's disease; sepsis; shock;  
 KW ischaemia-reperfusion injury; haematopoiesis; cancer; neuropathy;  
 KW transgenic animal; Alzheimer's disease; Parkinson's disease; burn;  
 KW amyotrophic lateral sclerosis; platelet disorder; thrombocytopenia;  
 KW ulcer; osteoporosis; bone degenerative disorder; periodontal disease;  
 KW gut protection; lung; liver fibrosis; immune deficiency; infection;  
 KW severe combined immunodeficiency; SCID; autoimmune disorder; allergy;  
 KW multiple sclerosis; rheumatoid arthritis; diabetes mellitus; asthma;  
 KW fertility; analgesic; pain; antigen.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200166689-A2.  
 XX  
 PD 13-SEP-2001.  
 XX  
 XX 05-MAR-2001; 2001WO-US04942.  
 XX  
 XX 07-MAR-2000; 2000US-0519705.  
 PR 19-MAY-2000; 2000US-0574454.  
 PR 17-JUN-2000; 2000US-0596193.  
 PR 14-JUL-2000; 2000US-0618847.  
 PR 19-SEP-2000; 2000US-0665363.  
 PR 20-OCT-2000; 2000US-0693267.  
 XX  
 XX (HYSE-) HYSEQ INC.  
 PA  
 XX

PI Tang YT, Liu C, Asundi V, Xu C, Wehrman T, Ren F, Ma Y, Zhou P;  
 PI Zhao QA, Yang Y, Drmanac RT, Zhang J, Chen R, Xue AJ, Wang J;  
 XX  
 DR WPI; 2001-589934/66.  
 DR N-PSDB; AAS45256.  
 XX  
 XX Novel polypeptides and nucleic acids obtained from cDNA libraries  
 PT prepared from various human tissues, for diagnosis and treatment of  
 PT cancer, neurological, inflammatory, and autoimmune disorders -  
 PS  
 XX Example 2; SEQ ID No 713; 107pp; English.  
 XX  
 CC The invention relates to novel isolated human secreted polypeptides (I)  
 CC and polynucleotides (II). (I) and (II) are useful for treating  
 CC inflammatory conditions such as arthritis, nephritis, Crohn's disease,  
 CC ischaemia-reperfusion injury, shock, sepsis, immune responses, and is  
 CC involved in increasing haematopoiesis, stem cell survival, bone growth  
 CC and remodeling. (II), (II) and modulators of (II) are useful for  
 CC prophylaxis or treatment of one or more cancers. (II) is also useful for  
 CC creating transgenic animals useful for studying the in vivo activities of  
 CC the polypeptide as well as for studying modulators of the polypeptides.  
 CC (I) induces the proliferation of neural cells and regeneration of nerve  
 CC and brain tissue and is useful for the treatment of central and  
 CC peripheral nervous system diseases and neuropathies, such as Alzheimer's,  
 CC Parkinson's disease, Huntington's disease, and amyotrophic lateral  
 CC sclerosis. In addition, (I) is involved in chemotactic or chemokinetic  
 CC activity, regulation of haematopoiesis and is useful for treating myeloid  
 CC or lymphoid cell disorders, platelet disorders such as thrombocytopenia  
 CC and for regeneration of bone, cartilage, tendon, ligament and/or nerve  
 CC tissue growth, and in tissue repair, healing of burns, incisions,  
 CC ulcers, for treating osteoporosis, osteoarthritis, bone degenerative  
 CC disorders, or periodontal disease. Furthermore, (I) is also useful for  
 CC gut protection or regeneration and treatment of lung or liver fibrosis,  
 CC reperfusion injury in various tissues, various immune deficiencies and  
 CC disorders including severe combined immunodeficiency (SCID), bacterial or  
 CC fungal infections, autoimmune disorders e.g. multiple sclerosis,  
 CC rheumatoid arthritis, diabetes mellitus, myasthenia gravis, allergic  
 CC reactions and conditions, such as asthma or other respiratory problems.  
 CC In addition, (I) affects biorhythms or circadian cycles of rhythms,  
 CC fertility, metabolism, catabolism, anabolism, storage or elimination of  
 CC dietary fat, lipid, protein, carbohydrate, vitamins, minerals, provides  
 CC analgesic effects or other pain reducing effects, immunoglobulin like  
 CC activity and can act as an antigen in a vaccine composition to raise an  
 CC immune response. AAU28020-AAU28395 represent novel human secreted protein  
 CC amino acid sequences of the invention.  
 XX  
 XX Sequence 127 AA;  
 CC  
 CC Query Match 92.8%; Score 64; DB 22; Length 127;  
 CC Best Local Similarity 92.9%; Pred. No. 0.012;  
 CC Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 CC  
 Qy 1 LILLRGLPGSGKTT 14  
 Db |||||  
 18 LILLRGLPGSGKTT 31  
 RESULT 5  
 AAB84326  
 ID AAB84326 standard; Protein; 243 AA.  
 XX  
 AC AAB84326;  
 XX  
 XX 22-AUG-2001 (first entry)  
 XX  
 XX Amino acid sequence of a human lyase and associated protein HLYAP-1.  
 DE  
 DE Human lyase and lyase associated protein; HLYAP; actinic keratosis;  
 KW cell proliferative disorder; arteriosclerosis; atherosclerosis; bursitis;  
 KW cirrhosis; hepatitis; psoriasis; mixed connective tissue disease;  
 KW myelofibrosis; cancer; reproductive disorder; neurological disorder;  
 KW cranial nerve disorder; spinal cord disease; muscular dystrophy;  
 KW neuromuscular disorder; peripheral nervous system disorder; anxiety;  
 KW



KW amnesia; schizophrenic disorder; inflammatory disorder;  
 KW Addison's disease; adult respiratory distress syndrome; allergy;  
 KW amyloidosis; anemia; asthma; autoimmune haemolytic anemia;  
 KW autoimmune thyroiditis; autoimmune polyendocrinopathy; atopic dermatitis;  
 KW Goodpasture's syndrome; gout; multiple sclerosis; osteoarthritis;  
 KW osteoporosis; psoriasis; rheumatoid arthritis; ulcerative colitis;  
 KW uveitis.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Modified-site 39  
 FT Modified-site /note= "potential phosphorylation site"  
 FT Modified-site 53  
 FT Modified-site /note= "potential phosphorylation site"  
 FT Modified-site 73  
 FT Modified-site /note= "potential phosphorylation site"  
 FT Modified-site 214  
 FT Modified-site /note= "potential phosphorylation site"  
 FT Modified-site 229  
 FT Modified-site /note= "potential glycosylation site"  
 XX WO200144445-A2.  
 XX  
 PD 21-JUN-2001.  
 XX  
 XX 13-DEC-2000; 2000WO-US33815.  
 XX  
 XX 16-DEC-1999; 99US-0172307.  
 XX  
 XX (INCY-) INCYTE GENOMICS INC.  
 XX  
 XX Yue H, Bandman O, Tang YT, Hillman JL, Lu DAM, Baughn MR;  
 XX WPI; 2001-390243/41.  
 DR N-PSDB; AAH24904.  
 XX  
 XX Novel human lyase proteins (HLVAP) useful for diagnosing, treating and  
 XX preventing neurological, reproductive, cell proliferative and  
 XX inflammatory disorders associated with abnormal expression of HLVAP -  
 XX  
 XX Claim 1; Page 88; 103pp; English.  
 XX  
 CC The present sequence represents a human lyase and lyase associated  
 CC protein, designated HLVAP. HLVAP is useful for diagnosing, treating or  
 CC preventing cell proliferative disorders such as actinic keratosis,  
 CC arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis,  
 CC psoriasis, mixed connective tissue disease (MCTD), myelofibrosis, cancer  
 CC reproductive disorders, neurological disorders, cranial nerve disorders,  
 CC spinal cord diseases, muscular dystrophy and other neuromuscular  
 CC disorders, peripheral nervous system disorders, anxiety, amnesia, and  
 CC schizophrenic disorders; inflammatory disorders such as acquired  
 CC immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory  
 CC distress syndrome, allergies, amyloidosis, anemia, asthma, autoimmune  
 CC haemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy  
 CC and Crohn's disease, atopic dermatitis, Goodpasture's syndrome, gout,  
 CC multiple sclerosis, osteoarthritis, osteoporosis, psoriasis, rheumatoid  
 CC arthritis or ulcerative colitis and uveitis.  
 XX  
 SQ Sequence 243 AA;  
 Query Match 92.8%; Score 64; DB 22; Length 243;  
 Best Local Similarity 92.9%; Pred. No. 0.024; 1; Indels 0; Gaps 0;  
 Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 LILLRLPGSGKTT 14  
 | | | | | | | | | |  
 Db 44 LVLLRLPGSGKTT 57  
 RESULT 6  
 AAB93057  
 ID AAB93057 standard; Protein; 659 AA.

XX AAB93057;  
 AC 26-JUN-2001 (first entry)  
 DT  
 XX Human protein sequence SEQ ID NO:11862.  
 DE  
 XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
 KW Homo sapiens.  
 OS  
 XX Homo sapiens.  
 PN EP1074617-A2.  
 XX  
 PD 07-FEB-2001.  
 XX  
 XX 28-JUL-2000; 2000EP-0116126.  
 XX  
 XX 29-JUL-1999; 99JP-0248036.  
 PR 27-AUG-1999; 99JP-0300253.  
 PR 11-JAN-2000; 2000JP-0118776.  
 PR 02-MAY-2000; 2000JP-0183767.  
 PR 09-JUN-2000; 2000JP-0241899.  
 XX  
 PA (HELI-) HELIX RES INST.  
 XX  
 XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX WPI; 2001-318749/34.  
 DR  
 XX Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length cDNAs -  
 XX  
 PS Claim 8; SEQ ID 11862; 2537pp + CD ROM; English.  
 XX  
 CC The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent polynucleotides, all of which are used in the exemplification  
 CC of the present invention.  
 XX  
 SQ Sequence 659 AA;  
 Query Match 85.5%; Score 59; DB 22; Length 659;  
 Best Local Similarity 84.6%; Pred. No. 0.41;  
 Matches 11; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 LILLRLPGSGKT 13  
 | | | | | | | | | |  
 Db 442 LVLLRLPGSGSKS 454  
 RESULT 7  
 ABB76271

ID ABB76271 standard; Protein; 394 AA.  
 XX  
 AC ABB76271;  
 XX  
 DT 12-AUG-2002 (first entry)  
 XX  
 DE Human IKK binding protein Y2H14.  
 XX  
 KW Y2H14; IKK binding protein; I kappa B kinase; cytostatic;  
 KW antiinflammatory; immunosuppressive; immunostimulant; human.  
 XX  
 OS Homo sapiens.  
 XX  
 PN US635722-B1.  
 XX  
 PD 02-APR-2002.  
 XX  
 PF 17-NOV-1998; 98US-0195188.  
 XX  
 PR 17-NOV-1998; 98US-0195188.  
 XX  
 PA (UUNY ) UNIV NEW YORK STATE RES.  
 XX  
 PI Marcu KB;  
 XX  
 DR WPI; 2002-442213/47.  
 DR N-PSDB; ABL57342.  
 XX  
 PT Polynucleotide encoding an I kappa B kinase binding protein Y2H14 and  
 PT the recombinant protein encoded for elucidating and controlling  
 PT pathways leading to inflammation and apoptosis -  
 XX  
 PS Claim 1; Column 11-14; 9pp; English.  
 XX  
 CC The present sequence is the protein sequence for a human I kappa B  
 CC kinase (IKK) binding protein, designated Y2H14. Y2H14 was shown,  
 CC using yeast two-hybrid screens, to bind to the C-terminal domain  
 CC regions of IKK-alpha and IKK-beta, specifically to the region of  
 CC the IKK proteins made up of the contiguous helix-loop-helix and  
 CC leucine zipper domains. The Y2H14 protein and other IKK binding  
 CC proteins are useful for elucidating and controlling pathways  
 CC leading to inflammation and apoptosis. They can also be used to  
 CC detect IKK complexes and modulate IKK activity in cells undergoing  
 CC signalling by inflammatory mediators, and to identify  
 CC therapeutically active agents that modulate the binding or  
 CC interaction of Y2H14 with IKK-alpha or -beta. Molecules that  
 CC prevent the formation of Y2H14/IKK complexes or inhibit the  
 CC dissociation of these complexes are useful for boosting the  
 CC immune system, or as immunosuppressives, or as antiinflammatory  
 CC agents.  
 XX  
 SQ Sequence 394 AA;  
 Query Match 81.2%; Score 56; DB 23; Length 394;  
 Best Local Similarity 69.2%; Pred. No. 0.73;  
 Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 LILLRGLPGSGKT 13  
 Db 84 VIMRGLPGSGKT 96  
 ::::|||||  
 RESULT 8  
 AAY72168  
 ID AAY72168 standard; Protein; 629 AA.  
 XX  
 AC AAY72168;  
 XX  
 DT 24-APR-2001 (first entry)  
 XX  
 DE Human RNA metabolism protein (RMEP-8).  
 XX  
 KW Human; RNA metabolism protein; RMEP; nootropic; neuroleptic; antiulcer;  
 KW

tranquilliser; antianaemic; antidiabetic; immunosuppressive; cytostatic;  
 antiasthmatic; antiinflammatory; anti-HIV; human immunodeficiency virus;  
 antiarthritic; antiarteriosclerotic; antiatherosclerotic; antiallergic;  
 antirheumatoid; antiparkinsonian; antithyroid; nephrotrophic; antigout;  
 thymometric; RMEP expression modulator; transgenic; spinal cord disease;  
 nervous system disorder; Alzheimer's disease; cancer; gene therapy;  
 neuromuscular disorder; hepatitis; cancer; cell proliferative disorder;  
 peripheral nervous system disorder; cirrhosis; cranial nerve disorder;  
 progressive neural autonomic nervous system disorder; Addison's disease;  
 amyotrophic lateral sclerosis; autoimmune disorder; drug screening.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200078952-A2.  
 XX  
 PD 28-DEC-2000.  
 XX  
 PF 15-JUN-2000; 2000WO-US16644.  
 XX  
 PR 17-JUN-1999; 99US-0139922.  
 XX  
 PA (INCY-) INCYTE GENOMICS INC.  
 PA (BAUG/) BAUGHN M R.  
 XX  
 PI Bandman O, Yue H, Lal P, Tang YT, Reddy R, Azimzai Y;  
 DR WPI; 2001-102723/11.  
 DR N-PSDB; AAD02351.  
 XX  
 PT New human RNA metabolism proteins (RMEP), useful for diagnosing,  
 PT treating, preventing nervous system, cell proliferative,  
 PT autoimmune/inflammatory disorders associated with abnormal expression  
 PT of RMEP -  
 XX  
 PS Claim 1; Page 89-90; 103pp; English.  
 XX  
 CC The present sequence is human RNA metabolism protein (RMEP-8) encoded by  
 CC a cDNA (Clone ID 3803409) obtained from BLADTUT03 cDNA library.  
 CC Agonists and antagonists of RMEP cDNA are useful for treating diseases or  
 CC conditions associated with altered expression of functional RMEP. RMEP  
 CC sequence or their mammalian homologues are useful for creating 'knock  
 CC out' or 'knock in' humanised animals or transgenic animals to model  
 CC human disease. RMEP sequence is useful in the diagnosis, prevention and  
 CC treatment of nervous system disorders e.g. Alzheimer's disease, Pick's  
 CC disease, Huntington's disease, Parkinson's disease, amyotrophic lateral  
 CC sclerosis, and other motor neuron disorders, progressive neural autonomic  
 CC nervous system disorders, cranial nerve disorders, spinal cord diseases,  
 CC muscular dystrophy and other neuromuscular disorders, peripheral nervous  
 CC system disorders, mental disorders including anxiety and schizophrenia,  
 CC amnesia etc, cell proliferative disorders e.g. actinic keratosis,  
 CC arteriosclerosis, atherosclerosis, cirrhosis, hepatitis, mixed connective  
 CC tissue disease (MCTD), cancers e.g. adenocarcinoma, leukemia, lymphoma,  
 CC melanoma etc., and autoimmune/inflammatory disorders such as acquired  
 CC immuno deficiency syndrome (AIDS), Addison's disease, allergies, anaemia,  
 CC asthma, diabetes mellitus, rheumatoid arthritis, Grave's disease and  
 CC autoimmune polyendocrinopathy-candidiasis-ectodermal dysplasia (APECED).  
 CC RMEP cDNA is useful for somatic or germline gene therapy. RMEP sequence  
 CC is useful several drug screening assays.  
 XX  
 SQ Sequence 629 AA;  
 Query Match 81.2%; Score 56; DB 22; Length 629;  
 Best Local Similarity 69.2%; Pred. No. 1.2;  
 Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 LILLRGLPGSGKT 13  
 .....  
 Db 319 VIMRGLPGSGKT 331

RESULT 9  
 AAU51368  
 ID AAU51368 standard; Protein; 241 AA.  
 XX  
 AC AAU51368;

XX  
 DT 27-FEB-2002 (first entry)  
 XX  
 DE Propionibacterium acnes immunogenic protein #12364.

XX KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
 KW dermatological; osteopathic; neuroprotectant.

OS Propionibacterium acnes.

XX WO2000181581-A2.

PN 01-NOV-2001.

PF 20-APR-2001; 2001WO-US12865.

XX 21-APR-2000; 2000US-199047P.

PR 02-JUN-2000; 2000US-208841P.

PR 07-JUL-2000; 2000US-216747P.

XX (CORI-) CORIXA CORP.

XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;

PI L'maisonneuve J, Zhang Y, Jen S, Carter D;

XX WPI; 2001-616774/71.

DR N-PSDB; AAS59551.

XX Propionibacterium acnes polypeptides and nucleic acids useful for

PT vaccinating against and diagnosing infections, especially useful for

PT treating acne vulgaris -

XX Claim 3; SEQ ID No 12563; 1069pp; English.

PS Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic

XX polypeptides. The proteins and their associated DNA sequences are used in

CC the treatment, prevention and diagnosis of medical conditions caused by

CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne

CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.

CC P. acnes is also involved in infections of bone, joints and the central

CC nervous system, however it is particularly involved in the inflammatory

CC lesions associated with acne vulgaris. A method for detecting the

CC presence or absence of P. acnes in a patient comprises contacting a

CC sample with a binding agent that binds to the proteins of the invention

CC and determining the amount of bound protein in the sample. The

CC polypeptides may be used as antigens in the production of antibodies

CC specific for P. acnes proteins. These antibodies can be used to

CC downregulate expression and activity of P. acnes polypeptides and

CC therefore treat P. acnes infections. The antibodies may also be used as

CC diagnostic agents for determining P. acnes presence, for example, by

CC enzyme linked immunosorbent assay (ELISA).

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 241 AA;

SQ

QY 1 LILLRGLPGSGKT 14  
 .....  
 Db 53 LILVRGVSQSGKT 66

RESULT 10  
 AAB41858  
 ID AAB41858 standard; Protein; 314 AA.  
 XX  
 AC AAB41858;

XX  
 DT 08-FEB-2001 (first entry)  
 XX  
 DE Human ORFX ORF1622 polypeptide sequence SEQ ID NO:3244.

XX KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
 KW vulnerable; antiparasitic; antiparkinsonian; neurotropic; neuroprotective;  
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
 KW antiviral; antibacterial; antifungal; antineumatic; antithyroid;  
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;  
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 KW cholesterol ester storage; systemic lupus erythematosus; infection;  
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
 KW thrombosis; contraceptive.

OS Homo sapiens.

PN WO200058473-A2.

XX 05-OCT-2000.

XX 31-MAR-2000; 2000WO-US08621.

XX 31-MAR-1999; 99US-0127607.

PR 02-APR-1999; 99US-0127636.

PR 05-APR-1999; 99US-0127728.

PR 30-MAR-2000; 2000US-0540763.

XX (CURA-) CURAGEN CORP.

XX Shimkets RA, Leach M;

XX WPI; 2000-602362/57.

DR N-PSDB; AAC76067.

XX Novel nucleic acids and peptides derived from open reading frame X,

PT useful for treating e.g. cancers, proliferative disorders,

PT neurodegenerative disorders and cardiovascular disease -

XX Claim 11; Page 2446-2447; 5507pp; English.

XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,

CC which represent the human ORFX open reading frames 1 to 3161. The ORFX

CC sequences have activities such as: cytostatic; hepatotropic; vulnerary;

CC antiparasitic; antiparkinsonian; neurotropic; neuroprotective;

CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;

CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;

CC antidiabetic; hypotensive; dermatological; immunosuppressive;

CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;

CC antithyroid; and antianemic. The sequences can be used for determining

CC the presence of or predisposition to, or preventing or treating

CC pathological conditions associated with an ORFX-associated disorder. The

CC nucleic acids can be used to express ORFX proteins in gene therapy

CC vectors. The proteins and nucleic acids may be used to treat cancers,

CC proliferative disorders, neurodegenerative disorders, osteoarthritis,

CC graft vs host disease, cardiovascular disease, diabetes mellitus,

CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus

Query Match 79.7%; Score 55; DB 22; Length 241;  
 Best Local Similarity 78.6%; Pred. NO. 0.64;  
 Matches 11; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

CC erythematous, severe combined immunodeficiency (SCID), AIDS, viral,  
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
CC coagulation; to inhibit thrombosis; and as a contraceptive.  
XX  
SQ Sequence 314 AA;  
Query Match 73.9%; Score 51; DB 21; Length 314;  
Best Local Similarity 76.9%; Pred. No. 3.6;  
Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 LILLRGLPGSGKT 13  
Db 201 VILMVGLPGSGKT 213  
RESULT 11  
ABG22415  
ID ABG22415 standard; Protein; 592 AA.  
XX  
AC ABG22415;  
XX  
DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #22406.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
DR N-PSDB; AAS86602.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity  
XX  
PS Claim 20; SEQ ID No 52774; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 592 AA;  
Query Match 73.9%; Score 51; DB 22; Length 592;  
Best Local Similarity 76.9%; Pred. No. 6.8;  
Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 LILLRGLPGSGKT 13  
Db 280 VILMVGLPGSGKT 292  
RESULT 12  
AAV52063  
ID AAV52063 standard; Protein; 20 AA.  
XX  
AC AAV52063;  
XX  
DT 12-JUL-2000 (first entry)  
XX  
DE P. horikoshi PHEN013 protein fragment #1.  
XX  
KW Polymerase; thermostable; nucleic acid elongation; sliding clamp protein;  
KW amplification; reverse transcription.  
XX  
OS Pyrococcus horikoshi.  
XX  
PN DE19840771-A1.  
XX  
PD 10-FEB-2000.  
XX  
PF 07-SEP-1998; 98DE-1040771.  
XX  
PR 06-AUG-1998; 98DE-1035653.  
XX  
PA (LION-) LION BIOSCIENCE AG.  
XX  
PI Voss H, Moeckel G, Kober I, Kilger C;  
XX WPI; 2000-207149/19.  
XX  
PT A thermostable in vitro polymerase complex for template-dependent  
PT elongation of nucleic acids in amplification or reverse transcription  
PT methods  
XX  
PS Disclosure; Figure 7; 152pp; German.  
XX  
CC This invention describes a novel thermostable in vitro complex for  
CC template-dependent elongation of nucleic acids which comprises a  
CC thermostable sliding clamp protein, which is connected with an  
CC elongation protein that shows thermostable polymerase activity. The  
CC thermostable in vitro accessory complex can be used to produce the  
CC thermostable in vitro complex, which is useful for template-dependent  
CC elongation of nucleic acids, e.g. for amplification or reverse  
CC transcription. This is useful for sequencing nucleic acids by the  
CC polymerase chain reaction or reverse transcriptase-polymerase chain  
CC reaction (RT-PCR). The complex can be used to mark nucleic acids.  
CC AAV52000-V52084 and AAV90752-W90799 represent proteins and protein  
CC fragments used to illustrate the method of the invention.  
XX  
SQ Sequence 20 AA;  
Query Match 72.5%; Score 50; DB 21; Length 20;  
Best Local Similarity 76.9%; Pred. No. 0.32;  
Matches 10; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
Qy 2 ILLRGLPGSGKT 14  
Db 3 LLLAGPPGSGKTT 15

## RESULT 13

AAW55652  
ID AAW55652 standard; Protein; 132 AA.

XX AC AAW55652;

XX DT 24-JUN-1998 (first entry)

XX DE H. pylori ORF 06p30223\_23476067\_cl\_119 cytoplasmic protein.

XX KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;  
KW identification; binding compound; bacteria; life cycle; activator;  
KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.

XX OS Helicobacter pylori.

XX PN WO9737044-Al.

XX PD 09-OCT-1997.

XX PF 27-MAR-1997; 97WO-US05223.

XX PR 06-DEC-1996; 96US-0761318.

XX PR 29-MAR-1996; 96US-0625811.

XX PR 02-APR-1996; 96US-0758731.

XX PR 25-OCT-1996; 96US-0736905.

XX PR 28-OCT-1996; 96US-0738859.

XX PA (ASTR ) ASTRA AB.

XX PI Alm RA, Smith D;

XX DR WPI; 1997-503122/46.

XX DR N-PSDB; AAV25061.

XX PT Helicobacter pylori nucleic acid sequences and encoded  
PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori  
PT infection and for diagnosis of H. pylori infection

XX PS Claims 14,96; Page 857; 1145pp; English.

XX This sequence represents a H. pylori cytoplasmic protein involved in  
CC nucleotide metabolism and transport. The protein may be used in a vaccine  
CC to prevent or treat H. pylori infection or to identify H. pylori  
CC polypeptide binding compounds, useful as potential H. pylori life cycle  
CC activators or inhibitors. The DNA and probes derived from it may be used  
CC for the identification of H. pylori in a sample and the diagnosis of H.  
CC pylori infection. Nucleic acid sequences complementary to the DNA act as  
CC antisense sequences and can be used to prevent the translation of  
CC H. pylori mRNA. Antibodies against the protein can be used in  
CC immunoassays to evaluate the abundance and distribution of H. pylori-  
CC specific antigens. The genomic sequence of H. pylori (ATCC 55679) was  
CC determined from overlapping contigs generated by mechanically shearing  
CC the bacterial DNA. The sequences were analysed for ORF of at least 180  
CC nucleotides, and the predicted coding regions defined by computer  
CC evaluation. To identify likely H. pylori antigens for vaccine  
CC development, the amino acid sequences predicted from various ORF were  
CC analysed for significant homology to other known or exported membrane  
CC proteins. Having identified and determined the sequences of interest,  
CC particular regions can be isolated from H. pylori by PCR amplification  
CC for recombinant polypeptide production, e.g. in E. coli hosts.

XX SQ Sequence 132 AA;

Query Match 72.5%; Score 50; DB 18; Length 132;

Best Local Similarity 71.4%; Pred. No. 2.2;

Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 LILLRGLPGSGKTT 14

DB 4 LFLIIGAPSGKTT 17

## RESULT 14

AAW55730  
ID AAW55730 standard; Protein; 191 AA.

XX AC AAW55730;

XX DT 13-JUL-1998 (first entry)

XX DE H. pylori ORF 06p30223\_23476067\_cl\_115 cytoplasmic protein.

XX KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;  
KW identification; binding compound; bacteria; life cycle; activator;  
KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis;  
KW nucleotide metabolism; transport.

XX OS Helicobacter pylori.

XX PN WO9737044-Al.

XX PD 09-OCT-1997.

XX PF 27-MAR-1997; 97WO-US05223.

XX PR 06-DEC-1996; 96US-0761318.

XX PR 29-MAR-1996; 96US-0625811.

XX PR 02-APR-1996; 96US-0758731.

XX PR 25-OCT-1996; 96US-0736905.

XX PR 28-OCT-1996; 96US-0738859.

XX PA (ASTR ) ASTRA AB.

XX PI Alm RA, Smith D;

XX DR WPI; 1997-503122/46.

XX DR N-PSDB; AAV25139.

XX PT Helicobacter pylori nucleic acid sequences and encoded  
PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori  
PT infection and for diagnosis of H. pylori infection

XX PS Claims 14,96; Page 1001,1002; 1145pp; English.

XX This sequence represents a Helicobacter pylori cytoplasmic protein  
CC involved in nucleotide metabolism and transport.  
CC The protein may be used in a vaccine to prevent or treat H. pylori  
CC infection or to identify H. pylori polypeptide binding compounds,  
CC useful as potential H. pylori life cycle activators or inhibitors. The  
CC DNA and probes derived from it may be used for the identification of  
CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic  
CC acid sequences complementary to the DNA act as antisense sequences and  
CC can be used to prevent the translation of H. pylori mRNA. Antibodies  
CC against the protein can be used in immunoassays to evaluate the abundance  
CC and distribution of H. pylori-specific antigens. The genomic sequence of  
CC H. pylori (ATCC 55679) was determined from overlapping contigs generated  
CC by mechanically shearing the bacterial DNA. The sequences were analysed  
CC for ORF of at least 180 nucleotides, and the predicted coding regions  
CC defined by computer evaluation. To identify likely H. pylori antigens for  
CC vaccine development, the amino acid sequences predicted from various ORF  
CC were analysed for significant homology to other known or exported  
CC membrane proteins. Having identified and determined the sequences of  
CC interest, particular regions can be isolated from H. pylori by PCR  
CC amplification for recombinant polypeptide production, e.g. in E. coli  
CC hosts.

XX SQ Sequence 191 AA;

Query Match 72.5%; Score 50; DB 18; Length 191;

Best Local Similarity 71.4%; Pred. No. 3.2;

Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 LILLRGLPGSGKTT 14

DB 4 LFLIIGAPSGKTT 17

Db 4 LFLIIGAPGSGKTT 17

RESULT 15  
AAW98583  
ID AAW98583 standard; Protein; 191 AA.  
XX  
AC AAW98583;  
XX  
XX 31-MAR-1999 (first entry)  
XX  
DE H. pylori GHPO 68 protein.  
XX  
KW GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;  
KW peptic ulcer disease.  
XX  
OS Helicobacter pylori.  
XX  
PN WO9843478-A1.  
XX  
PD 08-OCT-1998.  
XX  
PF 01-APR-1998; 98WO-US06371.  
XX  
PR 29-JUL-1997; 97US-0902615.  
PR 01-APR-1997; 97US-0833457.  
PR 24-JUN-1997; 97US-0881227.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
PA (INNR ) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.  
XX  
PI Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;  
XX  
PI WPI: 1998-542293/46.  
DR N-PSDB; AAX14302.  
XX  
PT New isolated Helicobacter polynucleotides - used to develop products  
PT for the diagnosis, prevention and treatment of Helicobacter  
PT infections and gastrointestinal diseases  
XX  
PS Claim 8; Page 1152; 2054pp; English.  
XX  
CC This sequence represents a Helicobacter pylori GHPO protein of the  
CC invention. The polypeptides can be used for preventing or treating  
CC Helicobacter infections, and gastroduodenal diseases associated with  
CC these infections, including acute, chronic, and atrophic gastritis, and  
CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be  
CC used for the production of antibodies. The products can also be used for  
CC detection and diagnosis.  
XX  
SQ Sequence 191 AA;

Query Match 72.5%; Score 50; DB 19; Length 191;  
Best Local Similarity 71.4%; Pred. No. 3.2;  
Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 LILLRGLPGSGKTT 14  
| | | | |  
Db 4 LFLIIGAPGSGKTT 17

Search completed: December 3, 2002, 14:28:34  
Job time : 31.4 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 3, 2002, 14:26:06 ; Search time 12.6 seconds  
(without alignments)  
106.816 Million cell updates/sec

Title: US-09-924-654-4\_COPY\_404\_417

Perfect score: 69

Sequence: 1 LILLRGLPGSGKTT 14

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	59	85.5	400	1 ESBOP3	2',3'-cyclic-nucle
2	59	85.5	404	2 A45670	2',3'-cyclic-nucle
3	59	85.5	420	1 ESM532	2',3'-cyclic-nucle
4	59	85.5	420	2 I56577	2',3'-cyclic-nucle
5	59	85.5	421	1 J21517	2',3'-cyclic-nucle
6	58	84.1	149	2 D82178	hypothetical prote
7	57	82.6	190	2 A02386	hypothetical prote
8	54	78.3	168	2 H75269	hypothetical prote
9	53	76.8	186	2 G72590	probable adenyllys
10	53	76.8	222	2 T34753	probable ATP/Gnp b
11	53	76.8	480	2 A03170	conserved hypothet
12	52	75.4	443	2 D69306	conserved hypothet
13	52	75.4	452	2 F71327	probable signal re
14	52	75.4	502	2 AF3222	conserved hypothet
15	52	75.4	980	2 S71090	peroxisome biogene
16	52	75.4	1069	2 T22138	hypothetical prote
17	52	75.4	1105	2 T22132	hypothetical prote
18	51	73.9	252	2 A64492	toxin sensitivity
19	51	73.9	1607	2 T04583	TMV resistance pro
20	50	72.5	133	2 D64609	conserved hypothet
21	50	72.5	133	2 H71905	hypothetical prote
22	50	72.5	162	2 T16163	hypothetical prote
23	50	72.5	191	2 D71946	adenylate kinase -
24	50	72.5	191	2 B64597	adenylate kinase -
25	50	72.5	192	2 F81412	thymidylate kinase
26	50	72.5	199	2 H84343	late competence pr
27	50	72.5	348	2 H84003	probable replicati
28	50	72.5	468	2 G71231	activator 1, repli
29	50	72.5	479	2 D75198	

30	50	72.5	481	2 F95402	conserved hypothet
31	50	72.5	508	2 C87702	signal recognition
32	50	72.5	806	2 S22765	heterogeneous ribo
33	50	72.5	1235	2 T06267	nematodes resistan
34	50	72.5	1257	2 T06269	root-knot nematode
35	49	71.0	162	2 C75497	hypothetical prote
36	49	71.0	187	2 F75072	adenylate kinase r
37	49	71.0	250	2 T21197	hypothetical prote
38	49	71.0	424	2 T20692	hypothetical prote
39	49	71.0	448	2 D71852	probable signal re
40	49	71.0	448	2 H64663	signal recognition
41	49	71.0	460	2 D84396	signal recognition
42	49	71.0	510	2 AC2404	hypothetical prote
43	49	71.0	521	2 H87111	signal recognition
44	49	71.0	525	2 D70747	probable fth prote
45	49	71.0	536	2 S46058	YTA1 protein-relat

#### ALIGNMENTS

##### RESULT 1

###### ESBOP3

N;Alternate names: CNPase; cyclic-CMP phosphodiesterase; nucleoside-2',3'-cyclic-phosphat  
C;Species: Bos primigenius taurus (cattle)  
C;Date: 31-Mar-1993 #sequence revision 31-Mar-1993 #text\_change 11-Jun-1999  
C;Accession: A26861; A27038; A29136; A43792  
R;Vogel, U.S.; Thompson, R.J.  
Nucleic Acids Res. 15, 7204, 1987  
A;Title: Nucleotide sequence of bovine retina 2',3'-cyclic nucleotide 3'-phosphohydroly  
A;Reference number: A26861; MUID:88015580; PMID:2821502  
A;Accession: A26861  
A;Molecule type: mRNA  
A;Residues: 1-400 <VOG>  
A;Cross-references: GB:Y00405; NID:9261; PIDN:CAA68466.1; PID:9262  
A;Experimental source: retina  
R;Vogel, U.S.; Thompson, R.J.  
FEBS Lett. 218, 261-265, 1987  
A;Title: Molecular cloning of the myelin specific enzyme 2',3'-cyclic-nucleotide 3'-phos  
A;Reference number: A27038; MUID:87247281; PMID:3036592  
A;Accession: A27038  
A;Molecule type: mRNA  
A;Residues: 1-400 <VO2>  
A;Cross-references: GB:M27606; NID:G162880; PIDN:AAA30457.1; PID:G162881  
A;Experimental source: brain  
R;Kurihara, T.; Fowler, A.V.; Takahashi, Y.  
J. Biol. Chem. 262, 3256-3261, 1987  
A;Title: cDNA cloning and amino acid sequence of bovine brain 2',3'-cyclic-nucleotide 3'  
A;Reference number: A29136; MUID:87137604; PMID:3029107  
A;Accession: A29136  
A;Molecule type: mRNA  
A;Residues: 1-317, 'A', 319-400 <KUR>  
A;Cross-references: GB:J02659; NID:G162876; PIDN:AAA30456.1; PID:G162877  
A;Note: this sequence has been corrected in reference A43792  
R;Kurihara, T.; Fowler, A.V.; Takahashi, Y.  
J. Biol. Chem. 262, 16754, 1987  
A;Title: cDNA cloning and amino acid sequence of bovine brain 2',3'-cyclic-nucleotide 3'  
A;Reference number: A43792  
A;Accession: A43792  
A;Molecule type: mRNA  
A;Residues: 316-320 <KU2>  
A;Cross-references: GB:J02659  
A;Note: this reference is a correction to reference A29136  
C;Superfamily: 2',3'-cyclic-nucleotide 3'-phosphodiesterase  
C;Keywords: phosphoric diester hydrolase

Query Match 85.5% Score 59; DB 1; Length 400;  
Best Local Similarity 78.6%; Pred. No. 0.26;  
Matches 11; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 1 LILLRGLPGSGKTT 14  
| : ||||| : |

Db 32 LFILRLPGSGKST 45

RESULT 2

A45670  
2',3'-cyclic-nucleotide 3'-phosphodiesterase (EC 3.1.4.37) 1 - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 22-Jun-1999  
C:Accession: A45670  
R:Bernier, L.; Alvarez, F.; Norgard, E.M.; Ratble, D.W.; Mentaberry, A.; Schembri, J.G.;  
J. Neurosci. 7, 2703-2710, 1987  
A:Title: Molecular cloning of a 2',3'-cyclic nucleotide 3'-phosphodiesterase: mRNAs with  
A:Reference number: A45670; MUID:87310616; PMID:3040924  
A:Accession: A45670  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-404 <ER>  
A:Cross-references: EMBL:M18630; NID:g203492; PIDN:AAA04939.1; PID:g203493  
A:Superfamily: 2',3'-cyclic-nucleotide 3'-phosphodiesterase  
C:Keywords: phosphoric diester hydrolase

Query Match 85.5%; Score 59; DB 2; Length 404;  
Best Local Similarity 78.6%; Pred. No. 0.27; Gaps 0;  
Matches 11; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LILRLPGSGKTT 14

| : ||||| : |||

Db 32 LFILRLPGSGKST 45

RESULT 3

ESMS32  
2',3'-cyclic-nucleotide 3'-phosphodiesterase (EC 3.1.4.37) 2, brain - mouse  
N:Alternate names: CNPase; cyclic-CMP phosphodiesterase; nucleoside-2',3'-cyclic-phospho  
N:Contains: 2',3'-cyclic-nucleotide 3'-phosphodiesterase 1  
C:Species: Mus musculus (house mouse)  
C:Date: 31-Mar-1993 #sequence\_revision 31-Mar-1993 #text\_change 11-Jun-1999  
C:Accession: A33788; A60056  
R:Kurihara, T.; Monoh, K.; Sakimura, K.; Takahashi, Y.  
Biochem. Biophys. Res. Commun. 170, 1074-1081, 1990  
A:Title: Alternative splicing of mouse brain 2',3'-cyclic-nucleotide 3'-phosphodiesteras  
A:Reference number: A33788; MUID:90358801; PMID:2167669  
A:Accession: A33788  
A:Molecule type: mRNA  
A:Residues: 1-420 <UR>  
A:Cross-references: GB:M31810  
A:Note: the authors translated the codon TGG for residue 191 as Gly  
R:Monoh, K.; Kurihara, T.; Sakimura, K.; Takahashi, Y.  
Biochem. Biophys. Res. Commun. 165, 1213-1220, 1989  
A:Title: Structure of mouse 2',3'-cyclic-nucleotide 3'-phosphodiesterase gene.  
A:Reference number: A33788; MUID:90121227; PMID:2558653  
A:Accession: A33788  
A:Molecule type: mRNA  
A:Residues: 21-420 <MUN>  
A:Cross-references: GB:M31810; NID:g192648; PIDN:AAA37429.1; PID:g309178  
A:Note: the authors translated the codon TGG for residue 191 as Gly  
R:Kurihara, T.; Takahashi, Y.; Fujita, N.; Sato, S.; Miyatake, T.  
Brain Res. Mol. Brain Res. 5, 247-250, 1989  
A:Title: Developmental expression of 2',3'-cyclic-nucleotide 3'-phosphodiesterase mRNA i  
A:Reference number: A60056; MUID:89260946; PMID:2542718  
A:Accession: A60056  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 82-227 <KU2>  
C:Comment: This protein is the third most abundant protein in myelin, after proteolipid  
C:Superfamily: 2',3'-cyclic-nucleotide 3'-phosphodiesterase  
C:Keywords: alternative splicing; brain; myelin; phosphoric diester hydrolase  
F:1-420/Product: 2',3'-cyclic-nucleotide 3'-phosphodiesterase 2 #status predicted <ALT2>  
F:21-420/Product: 2',3'-cyclic-nucleotide 3'-phosphodiesterase 1 #status predicted <ALT1

Query Match 85.5%; Score 59; DB 1; Length 420;  
Best Local Similarity 78.6%; Pred. No. 0.28;  
Matches 11; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LILRLPGSGKTT 14

| : ||||| : |||

Db 52 LFILRLPGSGKST 65

RESULT 4

I56577  
2',3'-cyclic-nucleotide 3'-phosphodiesterase (EC 3.1.4.37) - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 21-Feb-1997 #sequence\_revision 21-Feb-1997 #text\_change 22-Jun-1999  
C:Accession: I56577  
R:Gravel, M.; DeAngelis, D.; Braun, P.E.  
J. Neurosci. Res. 38, 243-247, 1994  
A:Title: Molecular cloning and characterization of rat brain 2',3'-cyclic nucleotide 3'  
A:Reference number: I56577; MUID:95018377; PMID:7932861  
A:Accession: I56577  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-420 <RES>  
A:Cross-references: GB:L16532; NID:g294526; PIDN:AAA64429.1; PID:g294527  
C:Genetics:  
C:Superfamily: 2',3'-cyclic-nucleotide 3'-phosphodiesterase  
C:Keywords: phosphoric diester hydrolase

Query Match 85.5%; Score 59; DB 2; Length 420;  
Best Local Similarity 78.6%; Pred. No. 0.28;  
Matches 11; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LILRLPGSGKTT 14

| : ||||| : |||

Db 52 LFILRLPGSGKST 65

RESULT 5

JC1517  
2',3'-cyclic-nucleotide 3'-phosphodiesterase (EC 3.1.4.37) long splice form - human  
N:Alternate names: 2',3'-cyclic nucleotide-3'-phosphohydrolase; CNPase; cyclic-CMP phos  
N:Contains: 2',3'-cyclic-nucleotide 3'-phosphodiesterase (EC 3.1.4.37) short splice for  
C:Species: Homo sapiens (man)  
C:Date: 14-Jul-1994 #sequence\_revision 14-Jul-1994 #text\_change 19-Jan-2001  
C:Accession: JC1517; JC1518; A48934; A27703; PC4368  
R:Monoh, K.; Kurihara, T.; Takahashi, Y.; Ichikawa, T.; Kumaniishi, T.; Hayaishi, S.; Min  
Gene 129, 297-301, 1993  
A:Title: Structure, expression and chromosomal localization of the gene encoding human  
A:Reference number: JC1517; MUID:93314977; PMID:8392017  
A:Accession: JC1517  
A:Molecule type: DNA  
A:Residues: 1-421 <MON>  
A:Cross-references: DDBJ:D13144; DDBJ:D13145; DDBJ:D13146; NID:g219398; PIDN:BAA39694.1  
A:Note: long form (isoform 1)  
A:Accession: JC1518  
A:Molecule type: DNA  
A:Residues: 21-421 <MO2>  
A:Cross-references: DDBJ:D13144; DDBJ:D13145; DDBJ:D13146; NID:g219398; PIDN:BAA02435.1  
A:Note: short form (isoform 2)  
R:Thompson, R.J.  
Biochem. Soc. Trans. 20, 621-626, 1992  
A:Title: 2',3'-cyclic nucleotide-3'-phosphohydrolase and signal transduction in central  
A:Reference number: A48934; MUID:93050745; PMID:1385234  
A:Accession: A48934  
A:Molecule type: DNA  
A:Residues: 1-194, Y, 196-343, Y, 345-421 <THO>  
A:Cross-references: GB:I546849; NID:g2558813; PIDN:AAB33928.1; PID:g2558814  
A:Note: this sequence is inconsistent with the nucleotide translation  
A:Note: sequence extracted from NCBI backbone (NCBIN:116965, NCBIN:116968, NCBIN:116970,  
R:Kurihara, T.; Takahashi, Y.; Nishiyama, A.; Kumaniishi, T.  
Biochem. Biophys. Res. Commun. 152, 837-842, 1988  
A:Title: CDNA cloning and amino acid sequence of human brain 2',3'-cyclic-nucleotide 3'  
A:Reference number: A27703; MUID:188209067; PMID:2835044  
A:Accession: A27703  
A:Molecule type: mRNA



A;Residues: 21-421 <KUR>  
A;Cross-references: GB:M19650; NID:G180686; PIDN:AAA35704.1; PID:G180687  
B;Stricker, R.; Kalbacher, H.; Reiser, G.  
Biochem. Biophys. Res. Commun. 237, 286-270, 1997  
A;Title: The epitope recognized by a monoclonal antibody in the myelin-associated protein F; 1-421/Product: 2',3'-cyclic-nucleotide 3'-phosphodiesterase short splice form #status F; 21-421/Product: 2',3'-cyclic-nucleotide 3'-phosphodiesterase short splice form #status F; 57-64/Region: nucleotide-binding motif A (P-loop)  
A;Accession: PC4368  
A;Molecule type: protein  
A;Residues: 21-58 <STR>  
C;Comment: This enzyme is a marker enzyme for oligodendroglia and myelin in the central C;Genetics:  
A;Gene: GDB:CNP  
A;Cross-references: GDB:128029; OMIM:123830  
A;Map position: 17q21-17q21  
A;Introns: 1/3; 226/1; 272/3  
C;Keywords: alternative initiators; alternative splicing; myelin; nucleotide binding; P-F; 1-421/Product: 2',3'-cyclic-nucleotide 3'-phosphodiesterase short splice form #status F; 21-421/Product: 2',3'-cyclic-nucleotide 3'-phosphodiesterase short splice form #status F; 57-64/Region: nucleotide-binding motif A (P-loop)  
Query Match 85.5%; Score 59; DB 1; Length 421;  
Best Local Similarity 78.6%; Pred. No. 0.28;  
Matches 11; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 LILLRGLPGSGKTT 14  
| : ||||| : ||  
Db 52 LFLIRGLPGSGKST 65  
| : ||||| : ||  
RESULT 6  
D82178  
hypothetical protein VC1610 [imported] - Vibrio cholerae (strain N16961 serogroup O1)  
C;Species: Vibrio cholerae  
C;Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001  
C;Accession: D82178  
R;Heidelber, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.; chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, F. I., R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
Nature 406, 477-483, 2000  
A;Title: DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
A;Reference number: A82035; MUID:20406833; PMID:10952301  
A;Accession: D82178  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-149 <HEI>  
A;Cross-references: GB:AB004239; GB:AB003852; NID:G9656130; PIDN:AAF94764.1; GSPDB:GN001  
A;Experimental source: serogroup O1; strain N16961; biotype El Tor  
C;Genetics:  
A;Gene: VC1610  
A;Map position: 1  
Query Match 84.1%; Score 58; DB 2; Length 149;  
Best Local Similarity 78.6%; Pred. No. 0.13;  
Matches 11; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 LILLRGLPGSGKTT 14  
| : ||||| : ||  
Db 30 LTLIRGLPGSGKST 43  
| : ||||| : ||  
RESULT 7  
AD2386  
hypothetical protein all4644 [imported] - Nostoc sp. (strain PCC 7120)  
C;Species: Nostoc sp.  
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C;Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002  
C;Accession: AD2386  
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Yatanabe, A.; Iriguchi, Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S. DNA Res. 8, 205-213, 2001  
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena PCC 7120  
A;Reference number: AB1807; MUID:21595285; PMID:11759840

A;Accession: AD2386  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-190 <KUR>  
A;Cross-references: GB:BA000019; PIDN:BAE76343.1; PID:G17133781; GSPDB:GN00179  
A;Experimental source: strain PCC 7120  
C;Genetics:  
A;Gene: all4644  
Query Match 82.6%; Score 57; DB 2; Length 190;  
Best Local Similarity 85.7%; Pred. No. 0.24;  
Matches 12; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 LILLRGLPGSGKTT 14  
| : ||||| : ||  
Db 11 LILLIGLPGSGKST 24  
| : ||||| : ||  
RESULT 8  
H75269  
hypothetical protein - Deinococcus radiodurans (strain R1)  
C;Species: Deinococcus radiodurans  
C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 31-Mar-2000  
C;Accession: H75269  
R;White, O.; Eisen, J.A.; Heidelber, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999  
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
A;Reference number: A75250; MUID:20036896; PMID:10567366  
A;Accession: H75269  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-168 <WHI>  
A;Cross-references: GB:AE002076; GB:AE000513; NID:G6460285; PIDN:AAF12017.1; PID:G64602;  
A;Experimental source: strain R1  
C;Genetics:  
A;Gene: DR2467  
A;Map position: 1  
Query Match 78.3%; Score 54; DB 2; Length 188;  
Best Local Similarity 78.6%; Pred. No. 0.59;  
Matches 11; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 LILLRGLPGSGKTT 14  
| : ||||| : ||  
Db 5 LFLVGLPGAGKTT 18  
| : ||||| : ||  
RESULT 9  
G72590  
probable adenyllylsulfate kinase APE1195 - Aeropyrum pernix (strain K1)  
C;Species: Aeropyrum pernix  
C;Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2000  
C;Accession: G72590  
R;Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takai awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; DNA Res. 6, 83-101, 1999  
A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix strain K1  
A;Reference number: A72450; MUID:99310339; PMID:10382966  
A;Accession: G72590  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-186 <KAW>  
A;Cross-references: DDBJ:AP000061; NID:G5104821; PIDN:BAA80181.1; PID:G5104867  
A;Experimental source: strain K1  
C;Genetics:  
A;Gene: APE1195  
C;Superfamily: adenyllylsulfate kinase; adenyllylsulfate kinase homology  
Query Match 76.8%; Score 53; DB 2; Length 186;  
Best Local Similarity 71.4%; Pred. No. 0.93;  
Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 LILLRGLPGSGKTT 14  
::| | | | | | | | | |  
Db 15 VVLTGLPGSGKTT 28

## RESULT 10

T34753

probable ATP/GTP binding protein - Streptomyces coelicolor

C;Species: Streptomyces coelicolor  
C;Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 05-Nov-1999

C;Accession: T34753

R;Murphy, L.; Harris, D.; Parkhill, J.; Barrrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, August 1998

A;Reference number: Z21556

A;Accession: T34753

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-222 <MUR>

A;Cross-references: EMBL:AL031184; PIDN:CAA20176.1; GSPDB:GN000070; SCOEDB:SC2A11.07

A;Experimental source: strain A3(2)

C;Genetics:

A;Gene: SCOEDB:SC2A11.07

Query Match 76.8%; Score 53; DB 2; Length 222;  
Best Local Similarity 64.3%; Pred. No. 1.1;  
Matches 9; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 LILLRGLPGSGKTT 14  
::| | | | | | | | | |  
Db 56 LWITGLPGSGKST 69

## RESULT 11

AC3170

conserved hypothetical protein Atu5086 [imported] - Agrobacterium tumefaciens (strain C58)

C;Species: Agrobacterium tumefaciens

C;Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 11-Jan-2002

C;Accession: AC3170

R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.

erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClell

; Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001

A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,

ster, E.W.

A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A;Reference number: AB2577; PMID:11743193

A;Accession: AC3170

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-480 <KUR>

A;Cross-references: GB:AE008687; PIDN:AAL45777.1; PID:gl7743512; GSPDB:GN00188

A;Experimental source: strain C58 (Dupont)

C;Genetics:

A;Gene: Atu5086

A;Genome: plasmid

Query Match 76.8%; Score 53; DB 2; Length 480;  
Best Local Similarity 71.4%; Pred. No. 2.5;  
Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 LILLRGLPGSGKTT 14  
| | | | | | | | | |  
Db 30 LYILQAGSGKTT 43

## RESULT 12

D69306

conserved hypothetical protein AF0452 - Archaeoglobus fulgidus

C;Species: Archaeoglobus fulgidus

C;Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 21-Jan-2000

C;Accession: D69306

R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson

; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.

Nature 390, 364-370, 1997

A;Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artlisch, P.; Kaine, B.P.; Sykes, S.

Smith, H.O.; Woese, C.R.; Venter, J.C.

A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archae

A;Reference number: A69250; MUID:98049343; PMID:9389475

A;Accession: D69306

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-443 <KLE>

A;Cross-references: GB:AE001073; GB:AE000782; NID:G2689396; PIDN:AAB90781.1; PID:G26501

C;Superfamily: Synchocystis conserved hypothetical protein sll1595

Query Match 75.4%; Score 52; DB 2; Length 443;  
Best Local Similarity 71.4%; Pred. No. 3.2;  
Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 LILLRGLPGSGKTT 14  
::| | | | | | | | | |  
Db 23 ILLIAGNPGSGKTT 36

## RESULT 13

F71327

probable signal recognition particle protein (ffh) - syphilis spirochete

C;Species: Treponema pallidum subsp. pallidum (syphilis spirochete)

C;Date: 24-Jul-1998 #sequence\_revision 24-Jul-1998 #text\_change 05-Nov-1999

C;Accession: F71327

R;Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwi

rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McD

they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.

Science 281, 375-388, 1998

A;Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.

A;Reference number: A71250; MUID:98332770; PMID:9665876

A;Accession: F71327

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-452 <COL>

A;Cross-references: GB:AE001219; GB:AE000520; NID:G3322693; PIDN:AAC65403.1; PID:G33226

A;Experimental source: strain Nichols

C;Genetics:

A;Gene: TP0416

C;Superfamily: signal recognition particle 54K protein

Query Match 75.4%; Score 52; DB 2; Length 452;  
Best Local Similarity 78.6%; Pred. No. 3.3;  
Matches 11; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 LILLRGLPGSGKTT 14  
::| | | | | | | | | |  
Db 103 IILLGLQSGKTT 116

## RESULT 14

AF3222

conserved hypothetical protein Atu5510 [imported] - Agrobacterium tumefaciens (strain C58)

C;Species: Agrobacterium tumefaciens

C;Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 11-Jan-2002

C;Accession: AF3222

R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.

erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClell

; Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001

A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,

ster, E.W.

A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A;Reference number: AB2577; PMID:11743193

A;Accession: AF3222

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-502 <KUR>

A;Cross-references: GB:AE008687; PIDN:AAL46196.1; PID:gl7743969; GSPDB:GN00188

A;Experimental source: strain C58 (Dupont)  
 C;Genetics:  
 A;Gene: Atu5510  
 A;Genome: plasmid

Query Match 75.4%; Score 52; DB 2; Length 502;  
 Best Local Similarity 71.4%; Pred. No. 3.6;  
 Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 LLLRGLPGSGKTT 14  
 : ||| |||||  
 Db 29 VLLGAPGSGKTT 42

## RESULT 15

S71090

peroxisome biogenesis disorder group 4 protein PXAA1 - human

C;Species: Homo sapiens (man)

C;Date: 14-Feb-1997 #sequence\_revision 13-Mar-1997 #text\_change 02-Feb-2001

C;Accession: S71090

R;Yahraus, T.; Braverman, N.; Dодt, G.; Kalish, J.E.; Morrell, J.C.; Moser, H.W.; Valle,

EMBO J. 15, 2914-2923, 1996

A;Title: The peroxisome biogenesis disorder group 4 gene, PXAA1, encodes a cytoplasmic

A;Reference number: S71090; MUID:96272151; PMID:8670792

A;Accession: S71090

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-980 &lt;YAH&gt;

A;Cross-references: EMBL:U56602; NID:gl354752; PIDN:AAC50655.1; FID:gl354753

C;Superfamily: Ftsh/SEC18/CDC48-type ATP-binding domain homology

C;Keywords: ATP; nucleotide binding; P-loop

F;470-477/Region: nucleotide-binding motif A (P-loop)

F;713-928/Domain: Ftsh/SEC18/CDC48-type ATP-binding domain homology &lt;VATP&gt;

F;744-751/Region: nucleotide-binding motif A (P-loop)

Query Match 75.4%; Score 52; DB 2; Length 980;  
 Best Local Similarity 76.9%; Pred. No. 7.2;  
 Matches 10; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 ILLRGLPGSGKTT 14  
 : ||| |||||  
 Db 466 VLLRGPFGCGKTT 478

Search completed: December 3, 2002, 14:29:08  
 Job time : 13.6 secs

GenCore.version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 3, 2002, 14:26:01 ; Search time 7 Seconds  
(without alignments)

82.953 Million cell updates/sec

Title: US-09-924-654-4\_COPY\_404\_417

Perfect score: 69

Sequence: 1 L1LLRCLPGSGKTT 14

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	59	85.5	400	1 CN37_BOVIN	P06623 bos taurus
2	59	85.5	404	1 CN37_RAT	P13233 rattus norv
3	59	85.5	420	1 CN37_MOUSE	P16330 mus musculu
4	59	85.5	421	1 CN37_HUMAN	P09543 homo sapien
5	56	81.2	1386	1 ZAP3_MOUSE	P09017 mus musculu
6	56	81.2	1822	1 ZAP3_HUMAN	P49750 homo sapien
7	53	76.8	186	1 CYSC_AERPE	P09066 aeropyrum p
8	53	76.8	978	1 PEX6_RAT	P54777 rattus norv
9	52	75.4	980	1 PEX6_HUMAN	Q13608 homo sapien
10	50	72.5	174	1 YG29_PYRAE	Q82wt4 pyrobaculum
11	50	72.5	191	1 KAD_HELPJ	Q92118 helicobacte
12	50	72.5	191	1 KAD_HELPJ	P56104 helicobacte
13	50	72.5	192	1 KAD_CAMJE	Q9phm8 campylobact
14	50	72.5	199	1 KTHV_HALN1	Q9hmv4 halobacteri
15	50	72.5	348	1 CMGA_BACHD	Q9K919 bacillus ha
16	50	72.5	824	1 ROU_HUMAN	Q00839 homo sapien
17	49	71.0	187	1 YH25_PYRAB	Q9v023 pyrococcus
18	49	71.0	407	1 YZRS5_CAEEL	Q19683 caenorhabdi
19	49	71.0	421	1 PSME_METAC	Q8ti98 methanosarc
20	49	71.0	424	1 YQ25_CAEEL	Q09535 caenorhabdi
21	49	71.0	448	1 SR54_HELPJ	Q92K62 helicobacte
22	49	71.0	448	1 SR54_HELPJ	P56005 helicobacte
23	49	71.0	460	1 SR54_HALN1	Q9hmv5 halobacteri
24	49	71.0	521	1 SR54_MYCLE	Q13013 mycobacteri
25	49	71.0	525	1 SR54_MYCTU	Q10963 mycobacteri
26	49	71.0	536	1 PCH2_YEAST	P38126 saccharomyc
27	48	69.6	169	1 AROK_CORGL	Q9x5d1 corynebacte
28	48	69.6	172	1 CGD7_HUMAN	Q9y3d8 homo sapien
29	48	69.6	207	1 URK_UREPA	Q9pqt9 ureaplasma
30	48	69.6	451	1 SR54_METJA	Q57565 methanococ
31	48	69.6	452	1 YP59_MYCTU	Q50739 mycobacteri
32	48	69.6	709	1 CDAT_PLAFA	P46468 plasmodium
33	47	68.1	40	1 KAD_STACA	P35141 staphylococ

## ALIGNMENTS

### RESULT 1

ID	CN37_BOVIN	STANDARD;	PRT;	400 AA.
AC	P06623;			
DT	01-JAN-1988 (Rel. 06, Created)			
DT	01-JUL-1993 (Rel. 26, Last sequence update)			
DT	30-MAY-2000 (Rel. 39, Last annotation update)			
DE	2',3'-cyclic nucleotide 3'-phosphodiesterase (EC 3.1.4.37) (CNP)			
DE	(CNPase).			
GN	CNP.			
OS	Bos taurus (Bovine).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;			
OC	Bovidae; Bovinae; Bos.			
OX	NCBI_TaxID=9913;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Brain;			
RX	MEDLINE=87137604; PubMed=3029107;			
RA	Kurihara T., Fowler A.V., Takahashi Y.;			
RT	"cDNA cloning and amino acid sequence of bovine brain 2',3'-cyclic-			
RT	nucleotide 3'-phosphodiesterase.";			
RL	J. Biol. Chem. 262:3256-3261(1987).			
RN	[2]			
RP	REVISION TO 318.			
RC	TISSUE=Brain;			
RA	Kurihara T., Fowler A.V., Takahashi Y.;			
RL	J. Biol. Chem. 262:16754-16754(1987).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Retina;			
RX	MEDLINE=88015580; PubMed=2821502;			
RA	Vogel U.S., Thompson R.J.;			
RT	"Nucleotide sequence of bovine retina 2',3'-cyclic nucleotide 3'-			
RT	phosphohydrolase.";			
RL	Nucleic Acids Res. 15:7204-7204(1987).			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=87247281; PubMed=3036592;			
RA	Vogel U.S., Thompson R.J.;			
RT	"Molecular cloning of the myelin specific enzyme 2',3'-cyclic-			
RT	nucleotide 3'-phosphohydrolase.";			
RL	FEBS Lett. 218:261-265(1987).			
CC	-!- CATALYTIC ACTIVITY: Nucleoside 2',3'-cyclic phosphate + H(2)O =			
CC	nucleoside 2'-phosphate.			
CC	-!- SUBCELLULAR LOCATION: FIRMLY BOUND TO MEMBRANE STRUCTURES OF BRAIN			
CC	WHITE MATTER.			
CC	-!- PTM: MET-1 MAY BE REMOVED AFTER TRANSLATION.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			

34	47	68.1	174	1	YA68_METTH	O27140 methanobact
35	47	68.1	186	1	YA08_PVRHO	O58736 pyrococcus
36	47	68.1	186	1	YC94_PYRFU	P58833 pyrococcus
37	47	68.1	188	1	YA41_SULSO	Q97290 sulfolobus
38	47	68.1	189	1	KTH1_SULSO	Q9uxg7 sulfolobus
39	47	68.1	204	1	ARF4_ARATH	Q9shu5 arabidopsis
40	47	68.1	205	1	KTHY_PVRHO	O59366 pyrococcus
41	47	68.1	212	1	KTHV_SANY3	O55593 synschocyst
42	47	68.1	215	1	KAD_STAAM	Q99840 staphylococ
43	47	68.1	241	1	VHBL_SHVX	Q04581 shallot vir
44	47	68.1	357	1	RUVB_STRCO	Q91291 streptomyce
45	47	68.1	433	1	SR54_ARCFU	O29633 archaeoglob

```
CC ENBL; J02659; AAA30456.1; ALT_SEQ.
DR ENBL; Y00405; CAA68466.1; -.
DR ENBL; M27606; AAA30457.1; -.
DR PIR; A26861; ESEOP3.
DR InterPro; IPR001230; Prenyl_site.
KW Hydrolase; Membrane; Brain.
SQ SEQUENCE 400 AA; 44875 MW; 3448FC367D647CF8 CRC64;

Query Match      85.5%; Score 59; DB 1; Length 400;
Best Local Similarity 78.6%; Pred. No. 0.05;
Matches 11; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LILLRGLPGSGKTT 14
   | : ||||| |||
Db 32 LFILRGLPGSGKST 45

RESULT 2
CN37 RAT
ID _CN37 RAT STANDARD; PRT; 404 AA.
AC P1323;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE 2',3'-cyclic nucleotide 3'-phosphodiesterase (EC 3.1.4.37) (CNP)
DE 2',3'-cyclic nucleotide 3'-phosphodiesterase (EC 3.1.4.37) (CNP)
DE (CNPase).
GN CNP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS CNPI AND CNPII).
RX MEDLINE=87310616; PubMed=3040924;
RA Bernier L., Alvarez F., Norgard E.M., Raible D.W., Mantaberry A.,
RA Schemmi J.G., Sabatini D.D., Colman D.R.;
RT "Molecular cloning of a 2',3'-cyclic nucleotide 3'-phosphodiesterase:
RT mRNAs with different 5' ends encode the same set of proteins in
RT nervous and lymphoid tissues";
RL J. Neurosci. 7:2703-2710(1987).
CC -1- CATALYTIC ACTIVITY: Nucleoside 2',3'-cyclic phosphate + H(2)O =
CC nucleoside 2'-phosphate.
CC -1- SUBCELLULAR LOCATION: FIRMLY BOUND TO MEMBRANE STRUCTURES OF BRAIN
CC WHITE MATTER.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR ENBL; M18630; AAA40939.1; -.
DR PIR; A45670; A45670.
DR InterPro; IPR001230; Prenyl_site.
KW Hydrolase; Membrane; Brain.
SQ SEQUENCE 404 AA; 45595 MW; 9C4CA888B6963BFF CRC64;

Query Match      85.5%; Score 59; DB 1; Length 404;
Best Local Similarity 78.6%; Pred. No. 0.05;
Matches 11; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LILLRGLPGSGKTT 14
   | : ||||| |||
Db 32 LFILRGLPGSGKST 45

RESULT 3
CN37 MOUSE
ID _CN37 MOUSE STANDARD; PRT; 420 AA.
AC P16330;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
```

```
DT 01-AUG-1990 (Rel. 15, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE 2',3'-cyclic nucleotide 3'-phosphodiesterase (EC 3.1.4.37) (CNP)
DE (CNPase).
GN CNP OR CNPI.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM CNPI).
RX MEDLINE=90121227; PubMed=2558653;
RA Monoh K., Kurihara T., Sakimura K., Takahashi Y.;
RT "Structure of mouse 2',3'-cyclic-nucleotide 3'-phosphodiesterase
RT gene.";
RL Biochem. Biophys. Res. Commun. 165:1213-1220(1989).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS CNPI AND CNPII).
RC STRAIN=DBA;
RX MEDLINE=90358801; PubMed=2167669;
RA Kurihara T., Monoh K., Sakimura K., Takahashi Y.;
RT "Alternative splicing of mouse brain 2',3'-cyclic-nucleotide 3'-
RT phosphodiesterase mRNA.";
RL Biochem. Biophys. Res. Commun. 170:1074-1081(1990).
CC -1- CATALYTIC ACTIVITY: Nucleoside 2',3'-cyclic phosphate + H(2)O =
CC nucleoside 2'-phosphate.
CC -1- SUBCELLULAR LOCATION: FIRMLY BOUND TO MEMBRANE STRUCTURES OF BRAIN
CC WHITE MATTER.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; CNPI/DNAI AND CNPII/DNAII
CC (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR ENBL; M31810; AAA37429.1; -.
DR ENBL; M58045; AAA37430.1; -.
DR ENBL; D38642; BAA07621.1; -.
DR ENBL; D38640; BAA07621.1; JOINED.
DR ENBL; D38641; BAA07621.1; JOINED.
DR ENBL; D38642; BAA07622.1; -.
DR ENBL; D38640; BAA07622.1; JOINED.
DR ENBL; D38641; BAA07622.1; JOINED.
DR PIR; A35708; ESMS32.
DR MGD; MGI:88437; Cnpl.
DR InterPro; IPR001230; Prenyl_site.
KW Hydrolase; Membrane; Brain; Alternative splicing.
FT VARSPPLIC 1 20 MISSING (IN ISOFORM CNPI).
FT CONFLICT 115 115 M -> I (IN REF. 1).
FT CONFLICT 136 136 L -> Y (IN REF. 1).
SQ SEQUENCE 420 AA; 47123 MW; DD96BFED47AD15D3 CRC64;

Query Match      85.5%; Score 59; DB 1; Length 420;
Best Local Similarity 78.6%; Pred. No. 0.052;
Matches 11; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LILLRGLPGSGKTT 14
   | : ||||| |||
Db 52 LFILRGLPGSGKST 65

RESULT 4
CN37 HUMAN
ID _CN37 HUMAN STANDARD; PRT; 421 AA.
AC P09543;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
```

DE 2',3'-cyclic nucleotide 3'-phosphodiesterase (EC 3.1.4.37) (CNP)  
DE (CNPase).  
GN Homo sapiens (Human).  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93050745; PubMed=1385234;  
RA Thompson R.J.;  
RT "2',3'-cyclic nucleotide 3'-phosphohydrolase and signal transduction.  
RT in central nervous system myelin.";  
RL Biochem. Soc. Trans. 20:621-626(1992).  
RN [2]  
RP SEQUENCE FROM N.A. (SHORT FORM).  
RC TISSUE=Brain;  
RX MEDLINE=88209067; PubMed=2835044;  
RA Kurihara T., Takahashi Y., Nishiyama A., Kumanishi T.;  
RT "cDNA cloning and amino acid sequence of human brain 2',3'-cyclic-  
RT nucleotide 3'-phosphodiesterase.";  
RL Biochem. Biophys. Res. Commun. 152:837-842(1988).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93314977; PubMed=8392017;  
RA Monoh K., Kurihara T., Takahashi Y., Ichikawa T., Kumanishi T.,  
RA Hayashi S., Minoshima S., Shimizu N.;  
RT "Structure, expression and chromosomal localization of the gene  
RT encoding human 2',3'-cyclic-nucleotide 3'-phosphodiesterase.";  
RL Gene 129:297-301(1993).  
RN [4]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93080285; PubMed=1360194;  
RA Douglas A.J., Fox M.F., Abbott C.M., Hinks L.J., Sharpe G.,  
RA Povey S., Thompson R.J.;  
RT "Structure and chromosomal localization of the human 2',3'-cyclic  
RT nucleotide 3'-phosphodiesterase gene.";  
RL Ann. Hum. Genet. 56:243-254(1992).  
RN [5]  
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
RC TISSUE=Brain, and Skin;  
RA Strausberg R.;  
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
CC -1- CATALYTIC ACTIVITY: Nucleoside 2',3'-cyclic phosphate + H(2)O =  
CC nucleoside 2'-phosphate.  
CC -1- SUBCELLULAR LOCATION: FIRMLY BOUND TO MEMBRANE STRUCTURES OF BRAIN  
CC WHITE MATTER.  
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: CNPI/DNAI AND CNPII/DNAII  
CC (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; S46849; AAB23928.2; JOINED.  
DR EMBL; S46843; AAB23928.2; JOINED.  
DR EMBL; S46845; AAB23928.2; JOINED.  
DR EMBL; S46846; AAB23928.2; JOINED.  
DR EMBL; M19650; AAA35704.1; -.  
DR EMBL; D13146; BAA39694.1; -.  
DR EMBL; D13144; BAA39694.1; JOINED.  
DR EMBL; D13145; BAA39694.1; JOINED.  
DR EMBL; D13146; BAA02435.1; -.  
DR EMBL; D13144; BAA02435.1; JOINED.  
DR EMBL; D13145; BAA02435.1; JOINED.  
DR EMBL; S50017; AAB24298.2; -.  
DR EMBL; S50013; AAB24298.2; JOINED.  
DR EMBL; S50014; AAB24298.2; JOINED.  
DR EMBL; S50016; AAB24298.2; JOINED.

DR EMBL; BC001362; AAH01362.1; -.  
DR EMBL; BC006392; AAH06392.1; -.  
DR EMBL; BC011046; AAH11046.1; -.  
DR EMBL; BC028040; AAH28040.1; -.  
DR PIR; A27703; A27703.  
DR PIR; JC1518; JC1518.  
DR Genew; HGNC:2158; CNP.  
DR MIM; 123830; -.  
DR InterPro; IPR001230; Prenyl\_site.  
KW Hydrolase; Membrane; Brain; Alternative splicing.  
FT VAAPSLP 1  
SQ SEQUENCE 421 AA; 47578 MW; CA6D0097DFD87255 CRC64;  
Query Match 85.5%; Score 59; DB 1; Length 421;  
Best Local Similarity 78.6%; Pred. No. 0.052;  
Matches 11; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
Oy 1 LILLRGLPGSGKTT 14  
Db 52 LFILRGLPGSGKST 65  
RESULT 5  
ZAP3 MOUSE STANDARD; PRT; 1386 AA.  
ID ZAP3\_MOUSE  
AC Q9R0I7;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Nuclear protein ZAP3.  
GN ZAP3 OR ZAP.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Fetal liver;  
RA Misawa K., Nosaka T., Kitamura T.;  
RT "A huge nuclear protein rich in proline similar to human hypothetical  
RT protein zap3 and zap13.";  
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; AB033168; BAA85182.1; -.  
DR MGD; MGI:1926195; Zap3.  
KW Nuclear protein.  
FT DOMAIN 15 204 PRO-RICH.  
FT DOMAIN 355 473 GLN-RICH.  
FT DOMAIN 925 1012 ARG-RICH.  
SQ SEQUENCE 1386 AA; 155130 MW; D862F9918ED221DF CRC64;  
Query Match 81.2%; Score 56; DB 1; Length 1386;  
Best Local Similarity 69.2%; Pred. No. 0.51;  
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
Oy 1 LILLRGLPGSGKT 13  
Db 1076 VVWRGLPGSGKT 1088  
RESULT 6  
ZAP3 HUMAN STANDARD; PRT; 1822 AA.  
ID ZAP3\_HUMAN  
AC P49750; P49752; Q9F1V7;

DT 01-OCT-1996 (Rel. 34, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Nuclear protein ZAP3 (ZAP133).  
GN ZAP3.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Rowen L., Madan A., Qin S., Abbasi N., Baradarani L., Birditt B.,  
RA Bloom S., Dors M., Dickhoff R., Fleetwood P., Harrison G., James R.,  
RA Kaur A., Madan A., Owen M.P., Ratcliffe A., Shafer T., Hood L.,  
RT "Sequencing of human chromosome 14q24.3 region."  
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE OF 539-847 AND 1397-1822 FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=95319502; PubMed=7596406;  
RA Sherrington R., Rogaeve E.I., Liang Y., Rogaeva E.A., Levesque G.,  
RA Ikeda M., Chi H., Lin C., Li G., Holman K., Tsuda T., Mar L.,  
RA Foncin J.-F., Bruni A.C., Montesi M.P., Sorbi S., Rainero I.,  
RA Pinessi L., Nee L., Chumakov I., Pollen D., Brookes A.,  
RA Sansau P., Polinsky R.J., Wasco W., da Silva H.A.R., Haines J.L.,  
RA Pericak-Vance M.A., Tanzi R.E., Roses A.D., Fraser P.E.,  
RA Rommens J.M., St George-Hyslop P.H.;  
RT "Cloning of a gene bearing missense mutations in early-onset familial  
RL Alzheimer's disease."  
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).  
CC -1- CAUTION: Ref.2 sequence differs from that shown due to a  
CC frameshift in position 1661.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AC007956; AAF61275.1; -  
DR EMBL; L40403; AAC42008.1; ALT\_FRAME.  
DR EMBL; L40400; AAC42006.1; -  
KW Nuclear protein.  
FT DOMAIN 15 205 PRO-RICH.  
FT DOMAIN 382 430 GLN-RICH.  
FT DOMAIN 807 1209 ARG-RICH.  
FT DOMAIN 1488 1577 ARG-RICH.  
FT CONFLICT 621 621 P -> S (IN REF. 2).  
FT CONFLICT 1404 1404 T -> I (IN REF. 2).  
FT CONFLICT 1821 1821 K -> E (IN REF. 2).  
SQ SEQUENCE 1822 AA; 204947 MW; 8B6CB83FE540C7D2 CRC64;  
  
Query Match 81.2%; Score 56; DB 1; Length 1822;  
Best Local Similarity 69.2%; Pred. No. 0.67;  
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 LILLRGLPGSGKT 13  
Db 1641 VVIMRGLPGSGKT 1653  
  
RESULT 7  
ID CYSC\_AERPE STANDARD; PRT; 186 AA.  
AC Q9YCR6;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Probable adenylylsulfate kinase (EC 2.7.1.25) (APS kinase) (Adenosine-  
DE 5'-phosphosulfate kinase) (ATP adenosine-5'-phosphosulfate 3'-

DE phosphotransferase).  
GN CYSC OR APE1195.  
OS Aeropyrum pernix.  
OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;  
OC Desulfurococaceae; Aeropyrum.  
OX NCBI\_TaxID=56636;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K1;  
RX MEDLINE=99110339; PubMed=10382966;  
RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,  
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,  
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,  
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,  
RA Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,  
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;  
RT "Complete genome sequence of an aerobic hyper-thermophilic  
RL crenarchaeon, Aeropyrum pernix K1";  
RL DNA Res. 6:83-101(1999).  
CC -1- FUNCTION: CATALYZES THE SYNTHESIS OF ACTIVATED SULFATE.  
CC -1- CATALYTIC ACTIVITY: ATP + adenylylsulfate = ADP + 3'-  
CC phosphoadenylylsulfate.  
CC -1- PATHWAY: SECOND STEP IN THE SULFATE ACTIVATION PATHWAY. INVOLVED  
CC IN THE ENERGY-GENERATING SULFUR OXIDATION PATHWAYS OF MANY  
CC CHENOAUTOTROPHIC MICROBES.  
CC -1- SIMILARITY: BELONGS TO THE APS KINASE FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AF000061; BAA80181.1; -  
DR InterPro; IPR002891; APS\_Kinase.  
DR Pfam; PF01583; APS\_kinase; 1.  
DR ProDom; PD002350; APS\_kinase; 1.  
DR TIGRFAMs; TIGR00455; apsk; 1.  
KW Transferase; Kinase; ATP-binding; Phosphorylation; Complete proteome.  
FT NPBIND 20 27 ATP (BY SIMILARITY).  
FT ACT\_SITE 94 94 FORMS THE PHOSPHOSERINE INTERMEDIATE (BY  
FT SIMILARITY).  
SQ SEQUENCE 186 AA; 21095 MW; 5E2FCC480C78AC35 CRC64;  
  
Query Match 76.8%; Score 53; DB 1; Length 186;  
Best Local Similarity 71.4%; Pred. No. 0.2;  
Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 LILLRGLPGSGKTT 14  
Db 15 VVWLTLPGSGKTT 28  
  
RESULT 8  
PEX6\_RAT  
ID PEX6\_RAT STANDARD; PRT; 978 AA.  
AC P54777; O55097;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Peroxisome assembly factor-2 (PAF-2) (Peroxisomal-type ATPase 1)  
DE (Peroxin-6).  
GN PEX6.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Fischer 344; TISSUE=Liver;  
RX MEDLINE=96083586; PubMed=7493019;

RA Tsukamoto T., Miura S., Nakai T., Yokota S., Shimozawa N.,  
RA Suzuki Y., Orii T., Fujiki Y., Sakai F., Bogaki A., Yasuno H.,  
RA Osumi T.,  
RT "Peroxisome assembly factor-2, a putative ATPase cloned by functional  
RT complementation on a peroxisome-deficient mammalian cell mutant.",  
RL Nat. Genet. 11:395-401(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Sprague-Dawley;  
RA Tsukamoto T., Hashiguchi N.;  
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: INVOLVED IN PEROXISOME BIOSYNTHESIS. REQUIRED FOR  
CC STABILITY OF THE PTS1 RECEPTOR.  
CC -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN, PEROXISOME.  
CC -!- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; D63673; BAA09824.1; -;  
CC EMBL; D89660; BAA24931.1; -;  
CC EMBL; D89657; BAA24931.1; JOINED.  
CC EMBL; D89658; BAA24931.1; JOINED.  
CC EMBL; D89659; BAA24931.1; JOINED.  
CC InterPro; IPR003593; AAA\_ATPase.  
CC InterPro; IPR003959; AAA\_ATPase\_cent.  
CC InterPro; IPR003960; AAA\_sub.  
CC Pfam; PF00004; AAA; 2.  
CC SMART; SM00382; AAA; 1.  
CC PROSITE; PS00674; AAA; 1.  
CC NP\_BIND; 470 477 ATP (POTENTIAL).  
CC NP\_BIND; 742 749 ATP (POTENTIAL).  
CC MUTAGEN; 746 776 K->A: NO LOSS OF FUNCTION.  
CC MUTAGEN; 748 748 K->A: LOSS OF FUNCTION.  
CC CONFLICT; 299 299 D->G (IN REF. 2).  
CC CONFLICT; 333 333 V->G (IN REF. 2).  
CC CONFLICT; 343 343 Q->R (IN REF. 2).  
CC CONFLICT; 546 546 R->C (IN REF. 2).  
CC CONFLICT; 546 546 R->C (IN REF. 2).  
CC SEQUENCE 978 AA; 104426 MW; F723193B7595EA97 CRC64;  
  
Query Match 76.8%; Score 53; DB 1; Length 978;  
Best Local Similarity 76.9%; Pred. No. 1;  
Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
  
QY 2 ILLRGLPGSGKTT 14  
Db 466 VLLQGPFGSGKTT 478  
  
RESULT 9  
PEX6 HUMAN  
ID PEX6 HUMAN STANDARD; PRT; 980 AA.  
AC Q13608; O99476;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Peroxisome assembly factor-2 (PAP-2) (Peroxisomal-type ATPase 1)  
DE (Peroxin-6).  
GN PEX6 OR PXAA1.  
OS Homo sapiens (Human)  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96272151; PubMed=8670792;  
RA Yahraus T., Braverman N., Dodi G., Kalish J.E., Morrell J.C.,

RA Moser H.W., Valle D., Gould S.J.;  
RT "The peroxisome biogenesis disorder group 4 gene, PXAA1, encodes a  
RT cytoplasmic ATPase required for stability of the PTS1 receptor.",  
RL EMBO J. 15:2914-2923(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97094178; PubMed=8940266;  
RA Fukuda S., Shimozawa N., Suzuki Y., Zhang Z., Tomatsu S.,  
RA Tsukamoto T., Hashiguchi N., Osumi T., Masuno M., Inaizumi K.,  
RA Kuroki Y., Fujiki Y., Orii T., Kondo N.;  
RT "Human peroxisome assembly factor-2 (PAP-2): a gene responsible for  
RT group C peroxisome biogenesis disorder in humans.",  
RL Am. J. Hum. Genet. 59:1210-1220(1996).  
RN [3]  
RP SEQUENCE FROM N.A. AND VARIANTS ZS GLN-812 AND TRP-812.  
RX MEDLINE=99335284; PubMed=10408779;  
RA Zhang Z., Suzuki Y., Shimozawa N., Fukuda S., Imamura A.,  
RA Tsukamoto T., Osumi T., Fujiki Y., Orii T., Wanders R.J.A.,  
RA Barth P.G., Moser H.W., Paton B.C., Beasley G.T., Kondo N.;  
RT "Genomic structure and identification of 11 novel mutations of the  
RT PEX6 peroxisome assembly factor-2' gene in patients with peroxisome  
RT biogenesis disorders.",  
RL Hum. Mutat. 13:487-496(1999).  
CC -!- FUNCTION: INVOLVED IN PEROXISOME BIOSYNTHESIS. REQUIRED FOR  
CC STABILITY OF THE PTS1 RECEPTOR.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -!- DISEASE: DEFECTS IN PEX6 ARE THE CAUSE OF PEROXISOME BIOGENESIS  
CC DISORDER GROUP 4 (PBD GROUP 4). PBD ARE A GROUP OF LETHAL  
CC INHERITED DISEASES. THEY ARE CLINICALLY DIVIDED INTO FOUR GROUPS;  
CC ZELLWEGER SYNDROME (ZS), NEONATAL ADRENOLEUKODYSTROPHY (NALD),  
CC INFANTILE REFSUM DISEASE (IRD), AND CLASSICAL RHIZOMELIC  
CC CHONDRODYSPLASIA PUNCTATA (RCDP). ZS IS THE MOST SEVERE PHENOTYPE  
CC OF THE PBD AND IS CHARACTERIZED BY THE ABSENCE OF PEROXISOMES AND  
CC BY MULTIPLE METABOLIC DEFECTS. PATIENTS WITH ZS HAVE SEVERE  
CC NEUROLOGICAL ABNORMALITIES, DYSMORPHIC FACIAL FEATURES,  
CC HEPATOMEGALY, AND MULTIPLE RENAL CYSTS; MOST DIE WITHIN 6 MONTHS  
CC OF BIRTH. NALD AND IRD ARE SIMILAR TO ZS BUT PROGRESSIVELY  
CC MILDER, WITH SOME IRD PATIENTS SURVIVING BEYOND THE THIRD DECADE  
CC OF LIFE. CLASSICAL RCDP, DISTINCT FROM THE ZS SPECTRUM, IS  
CC ASSOCIATED WITH A CHARACTERISTIC FACIAL APPEARANCE, RHIZOMELIA,  
CC ICHTHYOSIS, AND SEVERE NEUROLOGICAL IMPAIRMENT.  
CC -!- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U56602; AAC50655.1; -;  
CC EMBL; D83703; BAA12069.1; -;  
CC EMBL; AF108098; AAF62564.1; -;  
CC EMBL; AF108095; AAF62564.1; JOINED.  
CC EMBL; AF108096; AAF62564.1; JOINED.  
CC EMBL; AF108097; AAF62564.1; JOINED.  
CC Genew; HGNC:8859; PEX6.  
CC MIM; 601498; -;  
CC MIM; 214100; -;  
CC InterPro; IPR003593; AAA\_ATPase.  
CC InterPro; IPR003959; AAA\_ATPase\_cent.  
CC InterPro; IPR003960; AAA\_sub.  
CC Pfam; PF00004; AAA; 2.  
CC SMART; SM00382; AAA; 1.  
CC PROSITE; PS00674; AAA; 1.  
KW Peroxisome; ATP-binding; Repeat; Disease mutation.  
FT NP\_BIND 470 477 ATP (POTENTIAL).  
FT NP\_BIND 744 751 ATP (POTENTIAL).  
FT VARIANT 812 812 R -> Q (IN ZS).  
FT VARIANT 812 812 /FTId=VAR\_007918.  
FT VARIANT 812 812 R -> W (IN ZS; ATYPICAL).  
FT VARIANT 812 812 /FTId=VAR\_007919.



```

FT CONFLICT 77 77 S -> N (IN REF. 1).
SQ SEQUENCE 980 AA; 104060 MW; 0EC1C2A75CE0038F CRC64;

Query Match
Best Local Similarity 75.4%; Score 52; DB 1; Length 980;
Matches 10; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 ILLRGLPGSGKTT 14
:|||||
Db 466 VLLRGPGCGKTT 478

RESULT 10
YG29 PYRAE STANDARD; PRT; 174 AA.
AC Q82WT4;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Hypothetical protein PAE1629.
GN PAE1629.
OS Pyrobaculum aerophilum.
OC Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;
OC Thermoproteaceae; Pyrobaculum.
OX NCBI_TaxID=13773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IM2 / ATCC 51768 / DSM 7523;
RX PubMed=11792869;
RA Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,
RA Miller J.H.;
RT "Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum
aerophilum".
RL Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).
CC -1- SIMILARITY: BELONGS TO THE UPF0200 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE009831; AAL63615.1; -
DR InterPro; IPR000850; Adenylate_kin.
DR Pfam; PF00406; adenylatekinase; 1.
DR PRINTS; PR00094; ADENYLTKINASE.
DR PROSITE; PS00113; ADENYLATE_KINASE; 1.
KW Transferase; Kinase; ATP-binding; Complete proteome.
FT NP_BIND 9 16 ATP (BY SIMILARITY).
SQ SEQUENCE 174 AA; 19701 MW; 0AAGFC9CB005378 CRC64;

Query Match
Best Local Similarity 72.5%; Score 50; DB 1; Length 174;
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 LILLRGLPGSGKTT 14
:|||||
Db 4 VLVAGLPGSGKTT 17

RESULT 11
KAD HELPJ
ID_KAD HELPJ STANDARD; PRT; 191 AA.
AC Q9ZLL6;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Adenylate kinase (EC 2.7.4.3) (ATP-AMP transphosphorylase).
GN ADK OR JRP0561.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=85963;

[1]
RN SEQUENCE FROM N.A.
RP MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori.";
RL Nature 397:176-180(1999).
CC -1- FUNCTION: THIS SMALL UBIQUITOUS ENZYME IS ESSENTIAL FOR
CC MAINTENANCE AND CELL GROWTH.
CC -1- CATALYTIC ACTIVITY: ATP + AMP = ADP + ADP.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE ADENYLATE KINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001489; AAD06144.1; -
DR HSP; P27142; IZIN
DR InterPro; IPR000850; Adenylate_kin.
DR Pfam; PF00406; adenylatekinase; 1.
DR PRINTS; PR00094; ADENYLTKINASE.
DR PROSITE; PS00113; ADENYLATE_KINASE; 1.
KW Transferase; Kinase; ATP-binding; Complete proteome.
FT NP_BIND 9 17 ATP (BY SIMILARITY).
SQ SEQUENCE 191 AA; 21408 MW; E1CB400485FF0AE1 CRC64;

Query Match
Best Local Similarity 72.5%; Score 50; DB 1; Length 191;
Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 LILLRGLPGSGKTT 14
:|||||
Db 4 LFLIAGPGSGKTT 17

RESULT 12
KAD HELPJ
ID_KAD HELPJ STANDARD; PRT; 191 AA.
AC P56104;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Adenylate kinase (EC 2.7.4.3) (ATP-AMP transphosphorylase).
GN ADK OR HP0618.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khatak H.G., Glodek A.,
RA McKenney K., FitzGerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
```

```

RT pylori."
RL Nucleotide sequence of the food-borne pathogen Campylobacter jejuni
CC -!- FUNCTION: THIS SMALL UBIQUITOUS ENZYME IS ESSENTIAL FOR
CC MAINTENANCE AND CELL GROWTH.
CC -!- CATALYTIC ACTIVITY: ATP + AMP = ADP + ADP.
CC -!- SUBUNIT: MONOMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO THE ADENYLATE KINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000576; AAD07683.1; -
DR HSSP; P43188; 1ZAK.
DR TIGR; HP0618; -
DR InterPro; IPR000850; Adenylate_kin.
DR Pfam; PF00406; adenylatekinase; 1.
DR ProDom; PD000657; Adenylate_kin; 1.
DR PROSITE; PS00113; ADENYLATE_KINASE; 1.
DR TRANSFERASE; Kinase; ATP-binding; Complete proteome.
KW NP_BIND 9
FT ATP (BY SIMILARITY).
SQ SEQUENCE 191 AA; 21243 MW; EFB5C6492CE3F9 CRC64;

Query Match 72.5%; Score 50; DB 1; Length 191;
Best Local Similarity 71.4%; Pred. No. 0.59;
Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Oy 1 LILLRGLPGSGKTT 14
Db 4 LFLIIGAPGSGKTT 17

RESULT 13
KAD_CAMJE
ID KAD_CAMJE STANDARD; PRT; 192 AA.
AC Q9PHM8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Adenylate kinase (EC 2.7.4.3) (ATP-AMP transphosphorylase).
GN ADK OR CJ0639C.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=NCTC 11168; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Kettley J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Feitwell T., Holroyd S.,
RA Jagsis K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrell B.G.;
RA "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences."
RL Nature 403:665-668 (2000).
CC -!- FUNCTION: THIS SMALL UBIQUITOUS ENZYME IS ESSENTIAL FOR
CC MAINTENANCE AND CELL GROWTH.
CC -!- CATALYTIC ACTIVITY: ATP + AMP = ADP + ADP.
CC -!- SUBUNIT: MONOMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO THE ADENYLATE KINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; ALI39075; CAB75275.1; -
DR HSSP; P43188; 1ZAK.
DR InterPro; IPR000850; Adenylate_kin.
DR Pfam; PF00406; adenylatekinase; 1.
DR PRINTS; PR00094; ADENYLTKINASE.
DR PRODOM; PD000657; Adenylate_kin; 1.
DR PROSITE; PS00113; ADENYLATE_KINASE; 1.
DR TRANSFERASE; Kinase; ATP-binding; Complete proteome.
KW NP_BIND 9
FT ATP (BY SIMILARITY).
SQ SEQUENCE 192 AA; 21447 MW; D2AADDE3EAA59D1 CRC64;

Query Match 72.5%; Score 50; DB 1; Length 192;
Best Local Similarity 71.4%; Pred. No. 0.59;
Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Oy 1 LILLRGLPGSGKTT 14
Db 4 LFLIIGAPGSGKTT 17

RESULT 14
KTHY_HALN1
ID KTHY_HALN1 STANDARD; PRT; 199 AA.
AC Q9HN4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Probable thymidylate kinase (EC 2.7.4.9) (dTMP kinase).
GN TMK OR VNG1929G.
OS Halobacterium sp. (strain NRC-1).
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahaitas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogha J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlshocher M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;
RA "Genome sequence of Halobacterium species NRC-1."
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181 (2000).
CC -!- CATALYTIC ACTIVITY: ATP + thymidine 5'-phosphate = ADP + thymidine
CC 5'-diphosphate.
CC -!- PATHWAY: SYNTHESIS OF dTTP FROM dTMP.
CC -!- SIMILARITY: BELONGS TO THE THYMIDYLATE KINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE005090; AAG20116.1; -
DR HSSP; P37345; 5TMP.
DR InterPro; IPR000062; Thymidylate_kin.
DR Pfam; PF02223; Thymidylate_kin; 1.
DR TIGRFAMs; TIGR00041; dTMP_Kinase; 1.
DR PROSITE; PS01331; THYMIDYLATE_KINASE; FALSE NEG.
DR TRANSFERASE; Kinase; Nucleotide biosynthesis; ATP-binding;
KW Complete proteome.
FT NP_BIND 7
SQ SEQUENCE 199 AA; 21687 MW; 4C877D1C8B0D9F8 CRC64;

```

Query Match 72.5%; Score 50; DB 1; Length 199;  
 Best Local Similarity 71.4%; Pred. No. 0.62;  
 Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 LILLRGLPGSGKTT 14  
 | | | | |  
 Db 2 LVTLEGLDGGKTT 15

## RESULT 15

CMGA\_BACHD STANDARD; PRT; 348 AA.  
 AC Q9K919;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Comg operon protein i homolog.  
 GN COMGA OR BH2832.  
 OS Bacillus halodurans.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=86665;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C-125 / JCM 9153;  
 RX MEDLINE=20512582; PubMed=11058132;  
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,  
 RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,  
 RA Horikoshi K.;  
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus  
 RT halodurans and genomic sequence comparison with Bacillus subtilis.";  
 RL Nucleic Acids Res. 28:4317-4331(2000).  
 CC -!- FUNCTION: REQUIRED FOR UPTAKE OF DNA BY COMPETENT BACILLUS  
 CC SUBTILIS CELLS (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: INNER FACE OF THE MEMBRANE (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE PULE/OUTE/EXEE/XPSE/XCPR FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AP001516; BAB06551.1;  
 DR InterPro; IPR001482; GSPII\_E.  
 DR Pfam; PF00437; GSPII\_E; 1.  
 DR ProDom; PD000739; GSPII\_E; 1.  
 DR PROSITE; PS00662; T2SP\_E; 1.  
 KW Transport; ATP-binding; Complete proteome.  
 FT NP\_BIND 145 152 ATP (POTENTIAL).  
 SQ SEQUENCE 348 AA; 38599 MW; 1BAAD3D5D9B7C8B CRC64;

Query Match 72.5%; Score 50; DB 1; Length 348;  
 Best Local Similarity 78.6%; Pred. No. 1.1;  
 Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 LILLRGLPGSGKTT 14  
 | | | | |  
 Db 140 LILLTGATGSGKTT 153

Search completed: December 3, 2002, 14:26:27  
 Job time : 8 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model  
Run on: December 3, 2002, 14:36:02 ; Search time 23.8 Seconds  
(without alignments)  
121.204 Million cell updates/sec

Title: US-09-924-654-4\_COPY\_404\_417  
Perfect score: 69  
Sequence: 1 LILLRGLPGSGKTT 14

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL.21.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phage.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	69	100.0	575	11 Q91ZG4	Q91ZG4 mus musculus
2	69	100.0	583	4 Q92802	Q92802 homo sapien
3	64	92.8	145	11 Q91ZG7	Q91ZG7 mus musculus
4	64	92.8	173	4 Q8WTU5	Q8WTU5 mus musculus
5	64	92.8	238	11 Q91ZG6	Q91ZG6 mus musculus
6	64	92.8	403	13 Q57390	Q57390 rana catesb
7	60	87.0	122	16 Q92KU1	Q92KU1 rhizobium m
8	59	85.5	400	11 Q91V42	Q91V42 mus musculus
9	59	85.5	420	11 Q64575	Q64575 rattus norv
10	59	85.5	420	11 Q923F3	Q923F3 mus musculus
11	59	85.5	659	4 Q9NVK2	Q9NVK2 homo sapien
12	59	85.5	1399	4 Q9P2D4	Q9P2D4 homo sapien
13	58	84.1	149	16 Q9KRM6	Q9KRM6 vibrio chol
14	57	82.6	190	16 Q8YNC2	Q8YNC2 anabaena sp
15	56	81.2	386	4 Q96164	Q96164 homo sapien
16	55	79.7	383	10 Q93YT2	Q93YT2 arabidopsis

17	55	79.7	420	13	057389
18	55	79.7	644	10	Q9FM17
19	54	78.3	168	16	Q9RRM1
20	54	78.3	401	4	Q96PX9
21	54	78.3	403	17	Q8ZUG1
22	53	76.8	183	16	Q8RXY7
23	53	76.8	222	16	Q86568
24	53	76.8	504	16	Q8UKL5
25	53	76.8	981	11	Q991C9
26	52	75.4	443	17	Q29797
27	52	75.4	452	16	Q83431
28	52	75.4	458	2	Q9L9C9
29	52	75.4	502	16	Q8UJG7
30	52	75.4	738	4	Q8WYQ0
31	52	75.4	892	4	Q8WYQ1
32	52	75.4	980	4	Q8WYQ2
33	52	75.4	1069	5	Q9U3F2
34	52	75.4	1105	5	Q20371
35	51	73.9	171	16	Q8XMI7
36	51	73.9	252	17	Q58933
37	51	73.9	774	13	Q9YHY5
38	51	73.9	1607	10	Q65506
39	50	72.5	133	16	Q25418
40	50	72.5	133	16	Q9ZLC6
41	50	72.5	162	5	Q19804
42	50	72.5	421	5	Q8SX76
43	50	72.5	468	17	Q57853
44	50	72.5	479	17	Q9UWR2
45	50	72.5	479	17	Q9V2G3

ALIGNMENTS

RESULT 1

Q91ZG4 PRELIMINARY; PRT; 575 AA.  
AC Q91ZG4;  
DT 01-DEC-2001 (TRENBLrel. 19, Created)  
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)  
DE DM417G6.4 (novel protein).  
GN DM417G6.4.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Wall M.  
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; ALJ58892; CAC42205.1; -.  
SQ SEQUENCE 575 AA; 66250 MW; C41FAA8CC3EE4408 CRC64;

RESULT 2

Q92802 PRELIMINARY; PRT; 583 AA.  
ID Q92802;  
AC Q92802;  
DT 01-FEB-1997 (TRENBLrel. 02, Created)  
DT 01-FEB-1997 (TRENBLrel. 02, Last sequence update)  
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)  
DE Hypothetical 67.5 kDa protein.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Query Match 100.0%; Score 69; DB 11; Length 575;  
Best Local Similarity 100.0%; Pred No. 0.007; 0; Indels 0; Gaps 0;  
Matches 14; Conservative 0; Mismatches 0;  
QY 1 LILLRGLPGSGKTT 14  
DB 396 LILLRGLPGSGKTT 409  
QY 1 LILLRGLPGSGKTT 14  
DB 396 LILLRGLPGSGKTT 409

```
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96411650; PubMed=8812419;
RA Cough F.J., Rommens J.M., Neuhausen S.L., Belanger C., Dumont M.,
RA Kenneth A., Bell R., Berry S., Bogden R., Cannon-Albright L.,
RA Farid L., Frye C., Hattler T., Janacki T., Jiang P., Kehrer R.,
RA LeBlanc J.F., McArthur-Morrison J., McSweeney D., Miki Y., Peng Y.,
RA Samson C., Schroeder M., Snyder S.C., Stringfellow M., Stroup C.,
RA Swedlund B., Swensen J., Teng D., Thakur S., Tran T., Tranchant M.,
RA Welver-Pelchhaus J., Wong A.K.C., Shizuya H., Labrie F., Skolnick M.H.,
RA Goldgar D.E., Kamb A., Weber B.L., Tavtigian S.V., Sinaid J.,
RA "Generation of an integrated transcription map of the BRCA2 region on
RT chromosome 13q12-q13."
RT Chromosoma 36:86-99(1996);
RN [2]
RP SEQUENCE FROM N.A.
RA Rhodes S.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON;
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; U50532; AAC50875.1; -.
DR EMBL; AL049783; CAB42441.1; -.
DR EMBL; BC010643; AAH10643.1; -.
KW Hypothetical protein.
SQ SEQUENCE 583 AA; 67459 MW; 318DC4D81CD0FF2A CRC64;

Query Match 100.0%; Score 69; DB 4; Length 583;
Best Local Similarity 100.0%; Pred. No. 0.0071;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LILLRGLPGSGKTT 14
DB 404 LILLRGLPGSGKTT 417

RESULT 3
Q91ZG7 PRELIMINARY; PRT; 145 AA.
AC Q91ZG7;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE DM417G6.2.2 (novel protein (isoform 2)).
GN DM417G6.2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Wall M.;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL358892; CAC42203.1; -.
SQ SEQUENCE 145 AA; 17129 MW; 70023679FC802CBC CRC64;

Query Match 92.8%; Score 64; DB 11; Length 145;
Best Local Similarity 92.9%; Pred. No. 0.0098;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LILLRGLPGSGKTT 14
DB 42 LILLRGLPGSGKTT 55

RESULT 4
Q8WTU5 PRELIMINARY; PRT; 173 AA.
ID Q8WTU5
AC Q8WTU5;
```

```
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical 20.7 kDa protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Strausberg R.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC022188; AAH22188.1; -.
KW Hypothetical protein.
SQ SEQUENCE 173 AA; 20678 MW; E79510AF6736241E CRC64;

Query Match 92.8%; Score 64; DB 4; Length 173;
Best Local Similarity 92.9%; Pred. No. 0.012;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LILLRGLPGSGKTT 14
DB 44 LILLRGLPGSGKTT 57

RESULT 5
Q91ZG6 PRELIMINARY; PRT; 238 AA.
AC Q91ZG6;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE DM417G6.2.1 (novel protein (isoform 1)).
GN DM417G6.2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Wall M.;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL358892; CAC42202.1; -.
SQ SEQUENCE 238 AA; 28038 MW; 8E55135F6EC4E9F9 CRC64;

Query Match 92.8%; Score 64; DB 11; Length 238;
Best Local Similarity 92.9%; Pred. No. 0.017;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LILLRGLPGSGKTT 14
DB 42 LILLRGLPGSGKTT 55

RESULT 6
Q57390 PRELIMINARY; PRT; 403 AA.
AC Q57390;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE CNP1.
OS Rana catesbeiana (Bull frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Ranoidae; Ranidae; Rana.
OX NCBI_TaxID=8400;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=97465541; PubMed=9326261;
RA Kasama-Yoshida H., Tohyama Y., Kurihara T., Sakuma M., Kojima H.,
RA Tamai Y.;
```

RT "A comparative study of 2',3'-cyclic-nucleotide 3'-phosphodiesterase  
 RT in vertebrates: cDNA cloning and amino acid sequences for chicken and  
 RT bullfrog enzymes.";  
 RL J. Neurochem. 69:1335-1342(1997).  
 DR EMBL; AB002145; BAA24228.1; -.  
 DR InterPro; IPR001230; Prenyl site.  
 DR PROSITE; PS00294; PRENYLATION; UNKNOWN 1.  
 SQ SEQUENCE 403 AA; 45870 MW; D6B8068CC31D7429 CRC64;

Query Match 92.8%; Score 64; DB 13; Length 403;  
 Best Local Similarity 85.7%; Pred. No. 0.03;  
 Matches 12; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LILLRGLPGSGKTT 14  
 |:|||||:|  
 Db 33 LVLLRGLPGSGKST 46

RESULT 7  
 Q92KU1 PRELIMINARY; PRT; 122 AA.  
 ID Q92KU1  
 AC Q91V42  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Hypothetical protein R00172.  
 GN R00172 OR SMC02847.  
 OS Rhizobium meliloti (Sinorhizobium meliloti).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Rhizobiaceae; Sinorhizobium.  
 OX NCBI\_TaxID=382;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1021;  
 RX MEDLINE=21396507; PubMed=11481430;  
 RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,  
 Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,  
 Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,  
 Pohl T., Portetelie D., Puhler A., Purnelle B., Ramsperger U.,  
 Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.,  
 RT "Analysis of the chromosome sequence of the legume symbiont  
 RT Sinorhizobium meliloti strain 1021";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).  
 DR EMBL; AL591782; CAC41559.1; -.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 122 AA; 12663 MW; 256D30402E668FE7 CRC64;

Query Match 87.0%; Score 60; DB 16; Length 122;  
 Best Local Similarity 85.7%; Pred. No. 0.036;  
 Matches 12; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LILLRGLPGSGKTT 14  
 |:|||||:|  
 Db 2 LIIUGLPGSGKTT 15

RESULT 8  
 Q91V42 PRELIMINARY; PRT; 400 AA.  
 ID Q91V42  
 AC Q91V42  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Cyclic nucleotide phosphodiesterase 1.  
 GN CNP1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ILS, AND ISS;  
 RA Ehringer M.A., Thompson J., Conroy O., Xu Y., Yang F., Canniff J.,

RA Beeson M., Gordon L., Bennett B., Johnson T.E., Sikela J.M.;  
 RT "High-throughput Sequence Identification of Gene Coding Variants  
 RT within Alcohol-Related QTLs";  
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE ADENYLATE KINASE FAMILY.  
 DR EMBL; AF332056; AAK56085.1; -.  
 DR EMBL; AF332055; AAK56084.1; -.  
 DR MGD; MGI:88437; Cnpl.  
 DR InterPro; IPR001230; Prenyl site.  
 DR PROSITE; PS00294; PRENYLATION; UNKNOWN 1.  
 KW Kinase; Transferase.  
 SQ SEQUENCE 400 AA; 44554 MW; B974404499DE29BE CRC64;

Query Match 85.5%; Score 59; DB 11; Length 400;  
 Best Local Similarity 78.6%; Pred. No. 0.19;  
 Matches 11; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LILLRGLPGSGKTT 14  
 |:|||||:|  
 Db 32 LFILRGLPGSGKST 45

RESULT 9  
 Q64575 PRELIMINARY; PRT; 420 AA.  
 ID Q64575  
 AC Q64575  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE 2',3'-cyclic nucleotide 3'-phosphodiesterase (EC 3.1.3.37)  
 DE (Fragment).  
 GN CNP1.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=BRAIN;  
 RX MEDLINE=95018377; PubMed=7932861;  
 RA Gravel M., Deangelis D., Braun P.E.;  
 RT "Molecular cloning and characterization of rat brain 2',3'-cyclic  
 RT nucleotide 3'-phosphodiesterase isoform 2";  
 RL J. Neurosci. Res. 38:243-247(1994).  
 DR EMBL; L16532; AAG64429.1; -.  
 DR InterPro; IPR001230; Prenyl site.  
 DR PROSITE; PS00294; PRENYLATION; UNKNOWN 1.  
 KW Hydrolase.  
 FT NON TER 420 420  
 SQ SEQUENCE 420 AA; 47268 MW; 83001BBA2C057826 CRC64;

Query Match 85.5%; Score 59; DB 11; Length 420;  
 Best Local Similarity 78.6%; Pred. No. 0.2;  
 Matches 11; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LILLRGLPGSGKTT 14  
 |:|||||:|  
 Db 52 LFILRGLPGSGKST 65

RESULT 10  
 Q923F3 PRELIMINARY; PRT; 420 AA.  
 ID Q923F3  
 AC Q923F3  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Cyclic nucleotide phosphodiesterase 1.  
 GN CNP1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;

```
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=MAMMARY TUMOR;
RA Strausberg R.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE ADENYLATE KINASE FAMILY.
DR EMBL; BC005544; AAH05544.1; -
DR EMBL; BC021904; AAH21904.1; -
DR MGD; MGI:88437; Cnpl.
DR InterPro; IPR001230; Prenyl site.
DR PROSITE; PS00294; PRENYLATION; UNKNOWN_1.
KW Kinase; Transferase.
SQ SEQUENCE 420 AA; 47123 MW; 2FAFFDE5F0E99EAB CRC64;

Query Match 85.5%; Score 59; DB 11; Length 420;
Best Local Similarity 78.6%; Pred. No. 0.2;
Matches 11; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LILLRGLPGSGKTT 14
|:|||||:|
Db 52 LFLRGLPGSGKST 65

RESULT 11
Q9NVK2 PRELIMINARY; PRT; 659 AA.
AC Q9NVK2;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE CDNA FLJ10680 fis, clone NT2RP200573, weakly similar to 2',3'-cyclic
DE nucleotide 3'-phosphodiesterase (EC 3.1.4.37) (Fragment).
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Ishibashi T., Fujimori K.,
RA Tanai H., Kimata M., Watanabe M., Hiraoka S., Ishii S., Kawai Y.,
RA Saito K., Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K.,
RA Masuno Y., Kanehori K.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK001542; BAA91748.1; -
DR InterPro; IPR001899; Gram pos anchor.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
FT NON_TER 659
SQ SEQUENCE 659 AA; 73887 MW; 9974015C7D6EAC1F CRC64;

Query Match 85.5%; Score 59; DB 4; Length 659;
Best Local Similarity 84.6%; Pred. No. 0.33;
Matches 11; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LILLRGLPGSGKT 13
|:|||||:|
Db 442 LVLLRGLPGSGKS 454

RESULT 12
Q9P2D4 PRELIMINARY; PRT; 1399 AA.
AC Q9P2D4;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)
DE KIAA1413 protein (Fragment).
GN KIAA1413.
```

```
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Nagase T., Kikuno R., Ishikawa K., Hirose M., Ohara O.;
RL "Prediction of the coding sequences of unidentified human genes.XVI.
RT The complete sequences of 150 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 7:65-73 (2000).
DR EMBL; AB037834; BAA92651.1; -
DR InterPro; IPR002625; Smr/MutS2.
DR Pfam; PF01713; Smr; 1
DR SMART; SM00463; SMR; 1.
FT NON_TER 1
SQ SEQUENCE 1399 AA; 158124 MW; 10ADC0B597C22BE0 CRC64;

Query Match 85.5%; Score 59; DB 4; Length 1399;
Best Local Similarity 84.6%; Pred. No. 0.76;
Matches 11; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LILLRGLPGSGKT 13
|:|||||:|
Db 88 LVLLRGLPGSGKS 100

RESULT 13
Q9KRM6 PRELIMINARY; PRT; 149 AA.
AC Q9KRM6;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE Hypothetical protein VC1610.
GN VC1610.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OC NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EL TOR N16961 / SEROTYPE O1;
EX MEDLINE=20405833; Pubmed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae.";
RL Nature 406:477-483 (2000).
DR EMBL; AE004238; AAF94764.1; -
DR TIGR; VC1610; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 149 AA; 17450 MW; 45D56676A4FF8A9A CRC64;

Query Match 84.1%; Score 58; DB 16; Length 149;
Best Local Similarity 78.6%; Pred. No. 0.093;
Matches 11; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LILLRGLPGSGKTT 14
|:|||||:|
Db 30 LTLIRGLPGSGKST 43

RESULT 14
Q8YNC2 PRELIMINARY; PRT; 190 AA.
AC Q8YNC2;
```

```

DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein A114644.
GN ALL4644.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
EMBL; AP003597; BAB78343.1; -
DR InterPro; IPR000566; Lipoclin_cytFABP.
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 190 AA; 21447 MW; 639CF4D0B023183A CRC64;

Query Match 82.8%; Score 57; DB 16; Length 190;
Best Local Similarity 85.7%; Pred. No. 0.18;
Matches 12; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LILLRGLPGSGKTT 14
Db 11 LILLRGLPGSGKST 24

RESULT 15
Q96164 PRELIMINARY; PRT; 386 AA.
AC Q96164;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Unknown (Protein for IMAGE:4026352) (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SKIN;
RA Strausberg R.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC007792; AAH07792.1; -
FT NON TER 1
SQ SEQUENCE 386 AA; 44902 MW; 4E1F199C111E3338 CRC64;

Query Match 81.2%; Score 56; DB 4; Length 386;
Best Local Similarity 69.2%; Pred. No. 0.56;
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 LILLRGLPGSGKT 13
Db 76 VVIMRGLPGSGKT 88

```

Search completed: December 3, 2002, 14:27:26  
 Job time : 26.8 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 3, 2002, 14:26:06 ; Search time 10.7333 Seconds  
(without alignments)  
38.378 Million cell updates/sec

Title: US-09-924-654-4\_COPY\_404\_417

Perfect score: 69

Sequence: 1 LILLRGLPGSGKTT 14

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*

1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PGTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	69	100.0	583	2	US-08-616-392C-4
2	56	81.2	394	4	US-09-135-188-1
3	50	72.5	479	4	US-09-446-504-54
4	50	72.5	479	4	US-09-712-266-64
5	50	72.5	806	1	US-07-980-528-2
6	50	72.5	1255	3	US-08-947-823-3
7	50	72.5	1257	3	US-08-947-823-5
8	47	68.1	222	4	US-09-134-001C-3289
9	47	68.1	815	4	US-09-134-001C-3614
10	46	66.7	65	4	US-09-268-364-6
11	45	65.2	223	4	US-09-574-141A-69
12	45	65.2	229	4	US-09-134-001C-4113
13	45	65.2	318	2	US-08-671-947-2
14	45	65.2	670	2	US-08-366-547-2
15	45	65.2	670	4	US-09-292-858B-2
16	44	63.8	181	1	US-08-418-444A-7
17	44	63.8	233	4	US-09-161-662-2
18	44	63.8	233	4	US-08-993-825-2
19	44	63.8	485	4	US-09-134-001C-4294
20	44	63.8	523	2	US-08-923-772-2
21	44	63.8	523	4	US-09-385-283-2
22	44	63.8	652	4	US-08-987-123-2
23	43	62.3	154	4	US-09-134-001C-3873
24	43	62.3	221	4	US-09-291-170A-8
25	43	62.3	221	4	US-09-724-884-8
26	43	62.3	308	4	US-09-149-476-607
27	43	62.3	341	4	US-09-134-001C-3650

Query Match 100.0%; Score 69; DB 2; Length 583;  
Best Local Similarity 100.0%; Pred. No. 0.0022;

#### ALIGNMENTS

##### RESULT 1

US-08-616-392C-4

; Sequence 4, Application US/08616392C

; Patent No. 5998165

; GENERAL INFORMATION:

; APPLICANT: Goold, Richard D.

; APPLICANT: Akerblom, Ingrid E.

; APPLICANT: Seilhamer, Jeffrey

; APPLICANT: Coleman, Roger

; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES PANCA

; NUMBER OF SEQUENCES: 12

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Drive

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/616,392C

; FILING DATE: 15-MAR-1996

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/581,240

; FILING DATE: 29-DEC-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Billings, Lucy J.

; REGISTRATION NUMBER: 36,749

; REFERENCE/DOCKET NUMBER: PF-0052-1US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 650-855-0555

; TELEFAX: 650-845-4166

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 583 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; IMMEDIATE SOURCE:

; LIBRARY: hnt

; CLONE: 496071

; US-08-616-392C-4

Sequence 474, Appl  
Sequence 5, Appl  
Sequence 5, Appl  
Sequence 21, Appl  
Sequence 7, Appl  
Sequence 7, Appl  
Sequence 6, Appl  
Sequence 2, Appl  
Sequence 6, Appl  
Sequence 341, Appl  
Sequence 4, Appl  
Sequence 3224, Ap  
Sequence 13, Appl  
Sequence 4795, Ap  
Sequence 5436, Ap  
Sequence 1, Appl

399 4 US-09-149-476-474  
456 2 US-08-933-750C-5  
456 4 US-09-234-613-5  
456 4 US-09-268-364-21  
503 3 US-08-911-853-7  
503 4 US-09-479-409-7  
503 4 US-09-479-453-7  
806 4 US-08-999-774A-6  
2763 3 US-08-496-944-2  
28 2 US-08-724-354D-6  
28 3 US-09-270-984A-6  
239 4 US-08-858-207A-341  
264 4 US-09-035-382-4  
267 4 US-09-134-001C-3224  
317 1 US-08-221-750A-13  
338 4 US-09-134-001C-4795  
368 4 US-09-134-001C-5436  
406 2 US-08-222-719-1

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLLRGLPGSGKTT 14  
| | | | | | | | | |  
Db 404 LLLRGLPGSGKTT 417

## RESULT 2

US-09-195-188-1  
; Sequence 1, Application US/09195188  
; Patent No. 6365722  
; GENERAL INFORMATION:  
; APPLICANT: MARCU, KENNETH B.  
; TITLE OF INVENTION: Y2H14 AN IKK BINDING PROTEIN  
; FILE REFERENCE: 178-264  
; CURRENT APPLICATION NUMBER: US/09/195.188  
; CURRENT FILING DATE: 1998-11-17  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 394  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-195-188-1

Query Match 81.2%; Score 56; DB 4; Length 394;  
Best Local Similarity 69.2%; Pred. No. 0.18;  
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLLRGLPGSGKT 13  
| | | | | | | | | |  
Db 84 VVIMRGLPGSGKT 96

## RESULT 3

US-09-446-504-64  
; Sequence 64, Application US/09446504  
; Patent No. 6218150  
; GENERAL INFORMATION:  
; APPLICANT: UEMORI, Takashi  
; APPLICANT: SATO, Yoshimi  
; APPLICANT: FUJITA, Tomoko  
; APPLICANT: MIYAKE, Kazuo  
; APPLICANT: MUKAI, Hiroyuki  
; APPLICANT: ASADA, Kiyozo  
; APPLICANT: KATO, Ikunoshin  
; TITLE OF INVENTION: DNA POLYMERASE-RELATED FACTORS  
; FILE REFERENCE: 1422-408PCT  
; CURRENT APPLICATION NUMBER: US/09/446.504  
; CURRENT FILING DATE: 1999-12-23  
; PRIOR APPLICATION NUMBER: PCT/JP98/02845  
; PRIOR FILING DATE: 1998-06-24  
; PRIOR APPLICATION NUMBER: JP 9-187496  
; PRIOR FILING DATE: 1997-06-26  
; PRIOR APPLICATION NUMBER: JP 9-320692  
; PRIOR FILING DATE: 1997-11-27  
; NUMBER OF SEQ ID NOS: 92  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 64  
; LENGTH: 479  
; TYPE: PRT  
; ORGANISM: Pyrococcus furiosus

US-09-446-504-64  
Query Match 72.5%; Score 50; DB 4; Length 479;  
Best Local Similarity 76.9%; Pred. No. 2;  
Matches 10; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 ILLRGLPGSGKTT 14  
| | | | | | | | | |  
Db 46 LLLAGPFGSGKTT 58

Qy 2 ILLRGLPGSGKTT 14  
| | | | | | | | | |  
Db 46 LLLAGPFGSGKTT 58

## RESULT 4

US-09-712-266-64  
; Sequence 64, Application US/09712266  
; Patent No. 6331158  
; GENERAL INFORMATION:  
; APPLICANT: UEMORI, Takashi  
; APPLICANT: SATO, Yoshimi  
; APPLICANT: FUJITA, Tomoko  
; APPLICANT: MIYAKE, Kazuo  
; APPLICANT: MUKAI, Hiroyuki  
; APPLICANT: ASADA, Kiyozo  
; APPLICANT: KATO, Ikunoshin  
; TITLE OF INVENTION: DNA POLYMERASE-RELATED FACTORS  
; FILE REFERENCE: 1422-408PCT  
; CURRENT APPLICATION NUMBER: US/09/712.266  
; CURRENT FILING DATE: 2000-11-15  
; PRIOR APPLICATION NUMBER: US 09/446.504  
; PRIOR FILING DATE: 1999-12-23  
; PRIOR APPLICATION NUMBER: PCT/JP98/02845  
; PRIOR FILING DATE: 1998-06-24  
; PRIOR APPLICATION NUMBER: JP 9-187496  
; PRIOR FILING DATE: 1997-06-26  
; PRIOR APPLICATION NUMBER: JP 9-320692  
; PRIOR FILING DATE: 1997-11-27  
; NUMBER OF SEQ ID NOS: 92  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 64  
; LENGTH: 479  
; TYPE: PRT  
; ORGANISM: Pyrococcus furiosus  
US-09-712-266-64

Query Match 72.5%; Score 50; DB 4; Length 479;  
Best Local Similarity 76.9%; Pred. No. 2;  
Matches 10; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 ILLRGLPGSGKTT 14  
| | | | | | | | | |  
Db 46 LLLAGPFGSGKTT 58

## RESULT 5

US-07-980-528-2  
; Sequence 2, Application US/07980528  
; Patent No. 5457026  
; GENERAL INFORMATION:  
; APPLICANT: Dreyfuss, Gideon  
; APPLICANT: Kiledjian, Megerditch  
; APPLICANT: Portman, Douglas S.  
; TITLE OF INVENTION: METHODS OF PROMOTING INTERMOLECULAR  
; TITLE OF INVENTION: INTERACTIONS INVOLVING A NUCLEIC ACID AND MOLECULES  
; TITLE OF INVENTION: USEFUL  
; TITLE OF INVENTION: THEREFOR  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &  
; ADDRESSEE: No. 5457026ris  
; STREET: One Liberty Place 46th Floor  
; CITY: Philadelphia  
; STATE: Pennsylvania  
; COUNTRY: USA  
; ZIP: 19403

US-07-980-528-2  
Query Match 72.5%; Score 50; DB 4; Length 479;  
Best Local Similarity 76.9%; Pred. No. 2;  
Matches 10; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25; mb/MD  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/980.528  
FILING DATE: 19921120  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: DeLuca, Mark

```
;
;   REGISTRATION NUMBER: 33,229
;   REFERENCE/DOCKET NUMBER: UPN-0847
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (215) 568-3100
;   TELEFAX: (215) 568-3439
;   INFORMATION FOR SEQ ID NO: 2:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 806 amino acids
;   TYPE: AMINO ACID
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
;   US-07-980-528-2
;
; Query Match 72.5%; Score 50; DB 1; Length 806;
; Best Local Similarity 57.1%; Pred. No. 3.4;
; Matches 8; Conservative 5; Mismatches 1; Indels 0; Gaps 0;
;
Qy 1 LILLRGLPGSGKTT 14
Db 480 VVMIGLPGACKTT 493
;
; RESULT 6
; US-08-947-823-3
; Sequence 3, Application US/08947823
; Patent No. 6114605
; GENERAL INFORMATION:
; APPLICANT: Williamson, Valerie M.
; APPLICANT: Kaloshian, Isgouhi
; APPLICANT: Yaghoobi, Jafar
; APPLICANT: Bodeau, John
; APPLICANT: Milligan, Stephen
; TITLE OF INVENTION: Procedures and Materials for Confering
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/947,823
; FILING DATE: 09-OCT-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; ADDRESS: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/947,823
; FILING DATE: 09-OCT-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US97/18802
; FILING DATE: 09-OCT-1997
; APPLICATION NUMBER: US 60/028,191
; FILING DATE: 10-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Bastian, Kevin L.
; REGISTRATION NUMBER: 34,774
; REFERENCE/DOCKET NUMBER: 023070-070210US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1255 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-947-823-3
;
; Query Match 72.5%; Score 50; DB 3; Length 1255;
; Best Local Similarity 64.3%; Pred. No. 5.5;
; Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
;
Qy 1 LILLRGLPGSGKTT 14
Db 545 VISITGMPGSGKTT 558
;
; RESULT 8
; US-09-134-001C-3289
; Sequence 3289, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
```

```

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3289
; LENGTH: 222
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3289

Query Match      68.1%; Score 47; DB 4; Length 222;
Best Local Similarity 69.2%; Pred. No. 2,6;
Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 ILLRGLPGSGKTT 14
   |||||
Db 10 ILMGLPGAGKGT 22
   |||||

RESULT 9
US-09-134-001C-3614
; Sequence 3614, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3614
; LENGTH: 815
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3614

Query Match      68.1%; Score 47; DB 4; Length 815;
Best Local Similarity 64.3%; Pred. No. 11;
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 LILLRGLPGSGKTT 14
   |||||
Db 360 IMLTGGPGTKTT 373
   |||||

RESULT 10
US-09-268-364-6
; Sequence 6, Application US/09268364A
; Patent No. 6204063
; GENERAL INFORMATION:
; APPLICANT: Allen, Steve
; APPLICANT: Lee, Jian Ming
; APPLICANT: Lightner, Jonathan
; APPLICANT: Ogell, Joan
; TITLE OF INVENTION: PLANT GLYCOLYSIS AND RESPIRATION ENZYMES
; FILE REFERENCE: BB-1154
; CURRENT APPLICATION NUMBER: US/09/268,364A
; CURRENT FILING DATE: 1999-03-15
; EARLIER APPLICATION NUMBER: 60/079,387
; EARLIER FILING DATE: March 16, 1998
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: Microsoft Office 97

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3289
; LENGTH: 222
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3289

Query Match      68.1%; Score 47; DB 4; Length 222;
Best Local Similarity 69.2%; Pred. No. 2,6;
Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 ILLRGLPGSGKTT 14
   |||||
Db 10 ILMGLPGAGKGT 22
   |||||

RESULT 9
US-09-134-001C-3614
; Sequence 3614, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3614
; LENGTH: 815
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3614

Query Match      68.1%; Score 47; DB 4; Length 815;
Best Local Similarity 64.3%; Pred. No. 11;
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 LILLRGLPGSGKTT 14
   |||||
Db 360 IMLTGGPGTKTT 373
   |||||

RESULT 10
US-09-268-364-6
; Sequence 6, Application US/09268364A
; Patent No. 6204063
; GENERAL INFORMATION:
; APPLICANT: Allen, Steve
; APPLICANT: Lee, Jian Ming
; APPLICANT: Lightner, Jonathan
; APPLICANT: Ogell, Joan
; TITLE OF INVENTION: PLANT GLYCOLYSIS AND RESPIRATION ENZYMES
; FILE REFERENCE: BB-1154
; CURRENT APPLICATION NUMBER: US/09/268,364A
; CURRENT FILING DATE: 1999-03-15
; EARLIER APPLICATION NUMBER: 60/079,387
; EARLIER FILING DATE: March 16, 1998
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: Microsoft Office 97

```

```

; SEQ ID NO 6
; LENGTH: 65
; TYPE: PRT
; ORGANISM: Oryza sativa
US-09-268-364-6

Query Match      66.7%; Score 46; DB 4; Length 65;
Best Local Similarity 75.0%; Pred. No. 1;
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 3 LLRGLPGSGKTT 14
   |||||
Db 35 LLGPPGTGKTT 46
   |||||

RESULT 11
US-09-574-141A-69
; Sequence 69, Application US/09574141A
; Patent No. 6395490
; GENERAL INFORMATION:
; APPLICANT: Gonsalves, Dennis
; APPLICANT: Meng, Baozhong
; TITLE OF INVENTION: RUPESTRIS STEM PITTING ASSOCIATED VIRUS
; TITLE OF INVENTION: NUCLEIC ACIDS, PROTEINS, AND THEIR USES
; FILE REFERENCE: 07678/035005
; CURRENT APPLICATION NUMBER: US/09/574,141A
; CURRENT FILING DATE: 2000-05-18
; PRIOR APPLICATION NUMBER: 60/047,147
; PRIOR FILING DATE: 1997-05-20
; PRIOR APPLICATION NUMBER: 60/069,902
; PRIOR FILING DATE: 1997-12-17
; PRIOR APPLICATION NUMBER: 09/081,320
; PRIOR FILING DATE: 1998-05-19
; NUMBER OF SEQ ID NOS: 97
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 69
; LENGTH: 223
; TYPE: PRT
; ORGANISM: Rupestris stem pitting associated virus
US-09-574-141A-69

Query Match      65.2%; Score 45; DB 4; Length 223;
Best Local Similarity 61.5%; Pred. No. 5,5;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 2 ILLRGLPGSGKTT 14
   |||||
Db 24 IVHVPFGSKTT 36
   |||||

RESULT 12
US-09-134-001C-4113
; Sequence 4113, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4113
; LENGTH: 229
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4113

Query Match      65.2%; Score 45; DB 4; Length 229;

```

Best Local Similarity 69.2%; Pred. No. 5.7;  
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 ILLRGLPGSGKTT 14  
Db 42 IILNGASGSGKTT 54

RESULT 13  
US-08-671-947-2  
; Sequence 2, Application US/08671947  
; Patent No. 5827515  
; GENERAL INFORMATION:  
; APPLICANT: Shahabi Reynoso, Mitra  
; APPLICANT: Yamamoto, Takashi  
; APPLICANT: Cooper, Nicole H.  
; APPLICANT: Kalman, Sue S.  
; TITLE OF INVENTION: BACILLUS THURINGIENSIS SPORULATION GENE  
; NUMBER OF SEQUENCES: 37  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SANDOZ AGRO, INC.  
; STREET: 975 California Avenue  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; FILING DATE: US/08/671,947  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Marcus-Wyner, Lynn  
; REGISTRATION NUMBER: 34,859  
; REFERENCE/DOCKET NUMBER: 133-0724  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415/354-3588  
; TELEFAX: 415/857-1125  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 318 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-671-947-2

Query Match 65.2%; Score 45; DB 2; Length 318;  
Best Local Similarity 61.5%; Pred. No. 8.1;  
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 ILLRGLPGSGKTT 14  
Db 92 MLFGNFGTGKTT 104

RESULT 14  
US-08-366-547-2  
; Sequence 2, Application US/08366547  
; Patent No. 5843737  
; GENERAL INFORMATION:  
; APPLICANT: Chen, Lan Bo  
; APPLICANT: Bao, Shideng  
; TITLE OF INVENTION: A NEW CANCER ASSOCIATED GENE, PROTEIN  
; TITLE OF INVENTION: EXPRESSED THEREFROM AND USES THEREOF  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS  
; ADDRESSEE: CUSHMAN  
; STREET: 130 Water Street

; CITY: Boston  
; STATE: MA  
; COUNTRY: US  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/366,547  
; FILING DATE: 30-DEC-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Eisenstein, Ronald I.  
; REGISTRATION NUMBER: 30628  
; REFERENCE/DOCKET NUMBER: 45072  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 523-3400  
; TELEFAX: (617) 523-6440  
; TELEX: 200291 STRE UR  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 670 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-366-547-2

Query Match 65.2%; Score 45; DB 2; Length 670;  
Best Local Similarity 57.1%; Pred. No. 18;  
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 LILLRGLPGSGKTT 14  
Db 121 ILLITGPPGCGKTT 134

RESULT 15  
US-09-292-858B-2  
; Sequence 2, Application US/09292858B  
; Patent No. 6455681  
; GENERAL INFORMATION:  
; APPLICANT: Dean, Frank  
; APPLICANT: O'Donnell, Michael E.  
; TITLE OF INVENTION: DNA MOLECULES ENCODING SINGLE STRAND GAP RESPONSE  
; TITLE OF INVENTION: PROTEINS INVOLVED IN ACTIVATION OF A DNA REPAIR/CELL  
; TITLE OF INVENTION: CYCLE CHECKPOINT PATHWAY  
; FILE REFERENCE: 22221/1011  
; CURRENT APPLICATION NUMBER: US/09/292,858B  
; CURRENT FILING DATE: 1999-04-16  
; PRIOR APPLICATION NUMBER: 60/082,020  
; PRIOR FILING DATE: 1998-04-16  
; NUMBER OF SEQ ID NOS: 27  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 670  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-292-858B-2

Query Match 65.2%; Score 45; DB 4; Length 670;  
Best Local Similarity 57.1%; Pred. No. 18;  
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 LILLRGLPGSGKTT 14  
Db 121 ILLITGPPGCGKTT 134

Search completed: December 3, 2002, 14:29:37  
Job time : 11.7333 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 3, 2002, 14:26:02 ; Search time 33.6 Seconds  
(without alignments)  
63.453 Million cell updates/sec

Title: US-09-924-654-4\_COPY\_550\_565

Perfect score: 92

Sequence: 1 PSKSTORPPPPQGRQ 16

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_101002.\*

1:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21:	/SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22:	/SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23:	/SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	92	100.0	583	AAW26654	Human PANC1A assoc
2	92	100.0	594	AAW56501	Human prostate can
3	52	56.5	124	ABG23700	Novel human diagno
4	52	56.5	743	ABB61223	Drosophila melanog
5	51	55.4	64	AAAB34360	Gene 7 human secre
6	51	55.4	89	ABB11023	Human mini-collage
7	50	54.3	82	AAO01615	Human polypeptide
8	50	54.3	114	AAO10075	Human polypeptide
9	50	54.3	210	AAW69663	Human LAGE-1 clone
10	50	54.3	210	AAW70861	Human LAGE-1 unsp1

11	49	53.3	114	22	AAO08261	Human polypeptide
12	49	53.3	167	20	AAV48240	Human prostate can
13	49	53.3	301	22	ABB68068	Drosophila melanog
14	49	53.3	440	23	AAE23386	Human intracellular
15	49	53.3	442	22	ABG01174	Novel human diagno
16	49	53.3	708	23	ABB91504	Herbicidally activ
17	49	53.3	828	22	ABB68240	Drosophila melanog
18	49	53.3	901	23	AAO18086	Human N-methyl-D-a
19	49	53.3	903	22	AAW50440	Rice EDRI. Oryza
20	49	53.3	1043	22	AAU02199	Human glutamate re
21	48.5	52.7	83.	22	AAU12192	Human PRO1881 poly
22	48.5	52.7	84	21	AAU38533	Human secreted pro
23	48	52.2	114	21	AAG41081	Zea mays protein f
24	48	52.2	114	21	AAG41135	Zea mays protein f
25	48	52.2	114	21	AAG41187	Zea mays protein f
26	48	52.2	124	22	AAO01520	Human polypeptide
27	48	52.2	125	21	AAG41134	Zea mays protein f
28	48	52.2	125	22	ABBI7449	Human nervous syst
29	48	52.2	127	21	AAG41089	Zea mays protein f
30	48	52.2	135	21	AAG41175	Zea mays protein f
31	48	52.2	136	21	AAG41186	Zea mays protein f
32	48	52.2	146	21	AAG41174	Zea mays protein f
33	48	52.2	151	21	AAG41088	Zea mays protein f
34	48	52.2	1413	22	ABB60857	Drosophila melanog
35	48	52.2	1424	22	ABB60854	Drosophila melanog
36	48	52.2	2289	18	AAW14987	Protein derived fr
37	47	51.1	17	18	AAW39001	Peptide resembling
38	47	51.1	68	22	AAU5745	Protonibacterium
39	47	51.1	109	22	AAU14868	Novel bone marrow
40	47	51.1	147	22	AAO13067	Human polypeptide
41	47	51.1	259	20	AAV23989	TON1 protein encod
42	47	51.1	345	14	AAW42317	EBV VCA-p40 from O
43	47	51.1	647	22	ABB68375	Drosophila melanog
44	47	51.1	749	22	ABB62915	Drosophila melanog
45	47	51.1	1073	22	ABB70031	Drosophila melanog

#### ALIGNMENTS

RESULT 1  
AAW26654  
ID AAW26654 standard; Protein; 583 AA.  
XX  
AC AAW26654;  
XX  
DT 16-FEB-1998 (first entry)  
XX  
DE Human PANC1A associated with pancreatic cancer.  
XX  
KW PANC1A; pancreas cancer; diagnosis; therapy; human.  
XX  
OS Homo sapiens.  
XX  
PN WO9724437-A1.  
XX  
PD 10-JUL-1997.  
XX  
PF 18-DEC-1996; 96WO-US19981.  
XX  
PR 15-MAR-1996; 96US-0616392.  
PR 29-DEC-1995; 95US-0581240.  
XX  
(INCY-) INCYTE PHARM INC.  
XX  
PI Akerblom IE, Coleman R, Goold RD, Seilhamer JJ;  
XX  
DR WPI; 1997-363673/33.  
XX  
N-PSDB; AAT91010.  
XX  
PT DNA encoding PANC1A and PANC1B associated with pancreatic cancer -  
PT useful for diagnosis of pancreatic cancer and identifying risk of  
PT cancer

XX Disclosure; Page 46-47; 65pp; English.

XX This protein comprises human PANC1A, which is associated with

CC pancreatic cancer and possibly other cancers. Its sequence was

CC deduced from a full-length PANC1A cDNA clone (see AAT91010) derived

CC from human teratocarcinoma NT-2 cells. Specific fragments of

CC PANC1A (see AAT91008) and PANC1B (see AAT91009) cDNA are claimed.

CC Also claimed are: antisense molecules; expression vectors;

CC transformed host cells; and purified PANC1A and PANC1B

CC polypeptides. Diagnostic testing for the presence of PANC1A or

CC PANC1B in a sample provides a means for positively correlating

CC abnormal levels of PANC1A or PANC1B with pancreatic cancer

CC (claimed). The host cells can be cultured for the recombinant

CC production of PANC1A or PANC1B proteins (claimed). PANC1A and

CC PANC1B polypeptides can be used to screen for antagonists and

CC inhibitors, and to raise diagnostic antibodies.

XX

SQ Sequence 583 AA;

Query Match 100.0%; Score 92; DB 18; Length 583;

Best Local Similarity 100.0%; Pred. No. 0.00022;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PSHKSTQRPDPGGRQ 16

DB 550 PSHKSTQRPDPGGRQ 565

RESULT 2

AAB56501

ID AAB56501 standard; Protein; 594 AA.

AC AAB56501;

DT 13-MAR-2001 (first entry)

XX

DE Human prostate cancer antigen protein sequence SEQ ID NO:1079.

XX

KW Human; prostate cancer; prostate cancer antigen; detection; diagnosis;

KW neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;

KW vulnary; gastrointestinal; nephrotropic; antiinfective; gynaecological;

KW antibacterial; gene therapy; neural; immune; reproductive; renal;

KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;

KW wound; infectious disease.

XX

OS Homo sapiens.

XX

PN WO200055174-A1.

XX

PD 21-SEP-2000.

XX

PF 08-MAR-2000; 2000WO-US05988.

XX

PR 12-MAR-1999; 99US-0124270.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

PA (ROSE/) ROSEN C A.

XX

PI Rosen CA, Ruben SM;

XX

DR WPI; 2000-587513/55.

DR N-PSDB; AAF15704.

XX

PT Prostate cancer associated gene sequences, referred to as prostate

PT cancer antigens, useful for treatment, prevention, and diagnosis of

PT disorders such as prostate cancer -

XX

PS Claim 11; Page 1502-1504; 2338pp; English.

XX

CC AAF15566 to AAF16505 encode the human prostate cancer associated

CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.

CC The prostate cancer antigens can have neuroprotective, cytostatic,

CC

CC cardioactive, immunomodulatory, muscular, vulnary, gastrointestinal,

CC nephrotropic, antiinfective, gynaecological and antibacterial activities,

CC and can be used in gene therapy. The prostate cancer antigen

CC polynucleotides may be used for detection of prostate cancer, chromosome

CC identification, as chromosome markers, and for numerous other diagnostic

CC or research purposes. The prostate cancer antigens may be used to treat

CC disorders such as neural, immune, muscular, reproductive,

CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative

CC disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to

CC AAB57303 represent sequences used in the exemplification of the present

CC invention.

XX

SQ Sequence 594 AA;

Query Match 100.0%; Score 92; DB 21; Length 594;

Best Local Similarity 100.0%; Pred. No. 0.00022;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PSHKSTQRPDPGGRQ 16

DB 561 PSHKSTQRPDPGGRQ 576

RESULT 3

ABG23700

ID ABG23700 standard; Protein; 124 AA.

XX

AC ABG23700;

DT 18-FEB-2002 (first entry)

XX

DE Novel human diagnostic protein #23691.

XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder.

XX

OS Homo sapiens.

XX

PN WO200175067-A2.

XX

PD 11-OCT-2001.

XX

PF 30-MAR-2001; 2001WO-US08631.

XX

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Drmanac RT, Liu C, Tang YT;

XX

DR WPI; 2001-639362/73.

DR N-PSDB; AAS87887.

XX

PT New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess

PT biodiversity -

XX

PS Claim 20; SEQ ID No 54059; 103pp; English.

XX

CC The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The

CC polynucleotides are also used in diagnostics as expressed sequence tags

CC for identifying expressed genes. (I) is useful in gene therapy techniques

CC to restore normal activity of (II) or to treat disease states involving

CC (II). (II) is useful for generating antibodies against it, detecting or

CC quantitating a polypeptide in tissue, as molecular weight markers and as

CC a food supplement. (II) and its binding partners are useful in medical

CC imaging of sites expressing (II). (I) and (II) are useful for treating

CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 124 AA;

Query Match 56.5%; Score 52; DB 22; Length 124;  
Best Local Similarity 50.0%; Pred. No. 10;  
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 PSKSTORPPPPQGRQ 16

Db 81 PRRSPAEPPTGTE 96

RESULT 4

ABB61223  
ID ABB61223 standard; Protein; 743 AA.

XX  
AC ABB61223;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 10461.

DE Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX N-PSDB; ABL05326.

XX New isolated nucleic acid detection reagent for detecting 1000 or more  
XX genes from Drosophila and for elucidating cell signalling and cell-cell  
XX interactions -

XX Disclosure; SEQ ID NO 10461; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent  
XX capable of detecting 1000 or more genes from Drosophila. The invention is  
XX useful in developmental biology and in elucidating cell signalling and  
XX cell-cell interactions in higher eukaryotes for the development of  
XX insecticides, therapeutics and pharmaceutical drugs. The invention  
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
XX sequences (ABL01840-ABL16175) and the encoded proteins  
XX (ABB57737-ABB72072).

XX The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 743 AA;

Query Match 56.5%; Score 52; DB 22; Length 743;

Best Local Similarity 64.3%; Pred. No. 55;  
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 HKSTORPPPPQGRQ 16

Db 61 HKSNVPVPPPSRK 74

RESULT 5

AAB34360  
ID AAB34360 standard; Protein; 64 AA.

XX  
AC AAB34360;

XX 26-JAN-2001 (first entry)

XX Gene 7 human secreted protein homologous amino acid sequence #121.

XX Human; secreted protein; diagnosis; neuroprotective; cytostatic;  
KW cardioactive; immunomodulatory; muscular active general; vulnary;  
KW gastrointestinal; nephrotropic; antiinfective; gynaecological; and  
KW antibacterial; gene therapy; detection; cancer; chromosome marker;  
KW chromosome identification; neural disorder; immune disorder;  
KW muscular disorder; reproductive disorder; gastrointestinal disorder;  
KW pulmonary disorder; cardiovascular disorder; renal disorder;  
KW proliferative disorder; wound healing; infectious disease; preservative;  
KW food additive.

XX Mus musculus.

XX WO200056883-A1.

XX 28-SEP-2000.

XX 16-MAR-2000; 2000WO-US06822.

XX 23-MAR-1999; 99US-0126054.

XX 10-DEC-1999; 99US-0169916.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM, Komatsoulis G;

XX WPI; 2000-587666/55.

XX Human secreted proteins and gene sequences encoding them, useful for  
XX detecting, preventing, and treating disorders such as cancer,  
XX neurological disorders and immune system disorders -

XX Disclosure; Page 394; 429pp; English.

XX The polynucleotide sequences given in AAC59566 to AAC59614 encode the  
XX human secreted proteins given in AAB34299 to AAB34347. AAB34348 to  
XX AAB34437 represent human secreted polypeptide sequences and proteins  
XX homologous to them, which are given in the exemplification of the present  
XX invention. Human secreted proteins have activities based on the tissues  
XX and cells the genes are expressed in. Example of activities include:  
XX neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular  
XX active general; vulnary; gastrointestinal; nephrotropic;  
XX antiinfective; gynaecological; and antibacterial. The polynucleotides  
XX can be used for the detection of various disorders such as cancer,  
XX chromosome identification, as chromosome markers, and for numerous other  
XX diagnostic or research purposes. The secreted proteins can be used to  
XX treat disorders such as neural, immune, muscular, reproductive,  
XX gastrointestinal, pulmonary, cardiovascular, renal, and proliferative  
XX disorders, wound healing, and infectious diseases. The proteins can  
XX also be used as a food additive or preservative to increase or decrease  
XX storage capabilities. AAC59557 to AAC59565 and AAB34298 represent  
XX sequences used in the exemplification of the present invention.

XX Sequence 64 AA;

Query Match 55.4%; Score 51; DB 21; Length 64;



Best Local Similarity 56.2%; Pred. No. 7.5;  
Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 PSHKSTQRPDPGQ 16  
Db 48 PQRSQSPDPGQ 63

RESULT 6  
ID ABB11023 standard; peptide; 89 AA.  
XX ABB11023;  
AC ABB11023;  
XX  
DT 11-JAN-2002 (first entry)  
XX  
DE Human mini-collagen homologue, SEQ ID NO:1393.  
XX  
KW Human; cytokine; cell proliferation; cell differentiation; growth factor;  
KW haematopoiesis regulation; tissue growth; immunomodulator; activin;  
KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;  
KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;  
KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;  
KW chronic inflammatory condition; proliferative retinopathy;  
KW atherosclerosis; coronary heart disease; arterial ischaemia;  
KW bone disorder; osteoporosis; vascular growth disorder;  
KW tissue regeneration; wound healing; infection; immune disorder;  
KW cell culture; drug screening; gene therapy; antiinflammatory;  
KW antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;  
KW cytostatic; osteopathic; vasodilator; cardiac; virucide; antibacterial;  
KW antifungal; vulnery; antiulcer.  
XX  
OS Homo sapiens.  
XX  
XX WO200157188-A2.  
PN  
XX  
XX  
PD 09-AUG-2001.  
XX  
XX  
PF 05-FEB-2001; 2001WO-US03800.  
XX  
XX  
PR 03-FEB-2000; 2000US-0496914.  
PR 27-APR-2000; 2000US-0560875.  
XX  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Tang YT, Liu C, Drmanac RT;  
XX  
XX WPI; 2001-457740/49.  
DR N-PSDB; ABA08267.  
XX  
XX Human proteins and DNA encoding sequences useful for preventing,  
PT treating or ameliorating a medical condition in a mammalian subject  
PT e.g. arthritis and cancer -  
XX  
XX Claim 20; Page 137; 1963pp; English.  
PS  
XX Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and  
CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The  
CC invention also relates to vectors and recombinant host cells comprising a  
CC nucleotide of the invention, methods of producing the novel polypeptides,  
CC antibodies against the polypeptides, methods of detecting the nucleotides  
CC or polypeptides in a sample, and methods of identifying compounds which  
CC bind to polypeptides of the invention. Although novel, many of the  
CC polypeptides of the invention have homology to known proteins, thereby  
CC giving an insight into their probable biological activities, and hence  
CC potential therapeutic applications. The polypeptides of the invention may  
CC have various activities, including cytokine, cell proliferation or cell  
CC differentiation activities; stem cell growth factor activity;  
CC haematopoiesis regulatory activity; tissue growth activity;  
CC immunomodulatory activity; activin- or inhibin-related activities;  
CC chemotactic or chemokinetic activities; haemostatic, thrombotic or  
CC thrombolytic activities; receptor or ligand activities; or may be  
CC involved in oncogenesis, cancer cell proliferation or metastasis.

CC Depending on their biological activities, polypeptides and nucleotides of  
CC the invention are useful for preventing, treating or ameliorating medical  
CC conditions, e.g., by protein or gene therapy. Such conditions include  
CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell  
CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),  
CC proliferative retinopathy, atherosclerosis, coronary heart disease,  
CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal  
CC vascular growth. Polypeptides involved with tissue regeneration and  
CC repair (or nucleic acids encoding them) may be used to promote wound  
CC healing (e.g., of burns, incisions and ulcers), while those with  
CC immunomodulatory activities may be used in the treatment of viral,  
CC bacterial and fungal infections in addition to immune disorders.  
CC Polypeptides with growth factor activity may be used in cell cultures to  
CC promote cell growth. For example, such polypeptides may be used to  
CC manipulate stem cells in culture to give rise to neuroepithelial cells  
CC that can be used to augment or replace cells damaged by illness,  
CC autoimmune disease or accidental damage. The polypeptides and nucleotides  
CC may also be used in the diagnosis of the above conditions, and in drug  
CC screening techniques. The present sequence represents a novel human  
CC polypeptide of the invention.  
XX  
SQ Sequence 89 AA;  
Query Match 55.4%; Score 51; DB 22; Length 89;  
Best Local Similarity 61.5%; Pred. No. 10;  
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 PSHKSTQRPDPGQ 13  
Db 72 PSHAGLARPPPE 84

RESULT 7  
AA001615  
ID AA001615 standard; Protein; 82 AA.  
XX  
XX AA001615;  
XX  
XX 06-NOV-2001 (first entry)  
DT  
XX  
XX Human polypeptide SEQ ID NO 15507.  
DE  
XX  
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorders; arthritis; inflammation.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200164835-A2.  
FN  
XX  
XX 07-SEP-2001.  
PD  
XX  
XX 26-FEB-2001; 2001WO-US04927.  
PF  
XX  
XX 28-FEB-2000; 2000US-0515126.  
PR  
XX 18-MAY-2000; 2000US-0577409.  
PR  
XX (HYSE-) HYSEQ INC.  
PA  
XX Tang YT, Liu C, Drmanac RT;  
XX WPI; 2001-514838/56.  
DR N-PSDB; AAI81546.  
DR  
XX Isolated nucleic acids and polypeptides, useful for preventing  
PT diagnosing and treating e.g. leukaemia, inflammation and immune  
PT disorders -  
XX  
XX Claim 20; SEQ ID NO 15507; 1399pp + Sequence Listing; English.  
PS  
XX The invention relates to human polynucleotides (AAI79941-AAI93841) and  
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to

CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 82 AA;  
 Query Match 54.3%; Score 50; DB 22; Length 82;  
 Best Local Similarity 53.3%; Pred. No. 13;  
 Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;  
 QY 1 PSHKSTORPPPPQGR 15  
 Db 17 PGRKKTRPPPPGCK 31  
 RESULT 8  
 AAO10075  
 ID AAO10075 standard; Protein; 114 AA.  
 XX  
 AC AAO10075;  
 XX  
 DT 06-NOV-2001 (first entry)  
 XX  
 DE Human polypeptide SEQ ID NO 23967.  
 XX  
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KW nervous system disorders; arthritis; inflammation.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200164835-A2.  
 XX  
 PD 07-SEP-2001.  
 XX  
 PF 26-FEB-2001; 2001WO-US04927.  
 XX  
 PR 28-FEB-2000; 2000US-0515126.  
 PR 18-MAY-2000; 2000US-0577409.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Tang YT, Liu C, Drmanac RT;  
 XX  
 DR WPI; 2001-514838/56.  
 DR N-PSDB; AAI90006.  
 XX  
 PT Isolated nucleic acids and polypeptides, useful for preventing  
 PT diagnosing and treating e.g. leukaemia, inflammation and immune  
 PT disorders -  
 XX  
 PS Claim 20; SEQ ID NO 23967; 1399pp + Sequence Listing; English.  
 XX  
 CC The invention relates to human polynucleotides (AAI79941-AAI93841) and  
 CC the encoded proteins (AAO00010-AAO1910) that exhibit activity elating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and

CC inflammation.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 114 AA;  
 Query Match 54.3%; Score 50; DB 22; Length 114;  
 Best Local Similarity 56.2%; Pred. No. 18;  
 Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;  
 QY 1 PSHKSTORPPPPQGR 16  
 Db 81 PXTKIESRPPPPGCK 96  
 RESULT 9  
 AAW69663  
 ID AAW69663 standard; Protein; 210 AA.  
 XX  
 AC AAW69663;  
 XX  
 DT 27-OCT-1998 (first entry)  
 XX  
 DE Human LAGE-1 clone 1 protein sequence.  
 XX  
 KW Human; LAGE-1; tumour associated protein; LL-1; diagnosis; tumour.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9832855-A1.  
 XX  
 PD 30-JUL-1998.  
 XX  
 PF 27-JAN-1998; 98WO-US01445.  
 XX  
 PR 27-JAN-1997; 97US-0791495.  
 XX  
 PA (LUDW-) LUDWIG INST CANCER RES.  
 XX  
 PI Boon-Falleur T, De Smet C, Godelaine D, Lethe B;  
 PI Lucas S;  
 XX  
 DR WPI; 1998-427951/36.  
 DR N-PSDB; AAV50346.  
 XX  
 PF New isolated LAGE-1 tumour associated nucleic acids - used to  
 PT develop products for the diagnosis and treatment of LAGE-1  
 PT associated disorders, particularly tumours  
 XX  
 PS Claim 21; Page 52-53; 73pp; English.  
 XX  
 CC The present sequence represents LAGE-1 tumour associated protein (TAP).  
 CC The present invention also describes: (1) a method for treating a  
 CC subject with a disorder characterised by expression of a LAGE-1 nucleic  
 CC acid molecule or an expression product, comprising administering to the  
 CC subject autologous cytolytic T cells to ameliorate the disorder, where  
 CC the cytolytic T cells are specific for complexes of an HLA molecule and  
 CC a LAGE-1 TAP or an immunogenic fragment; (2) a method for treating a  
 CC subject with a disorder characterised by expression of a LAGE-1 nucleic  
 CC acid molecule or an expression product, comprising administering a  
 CC LAGE-1 TAP or an immunogenic fragment to ameliorate the disorder; and  
 CC (3) a method for selectively enriching a population of T cells with  
 CC cytolytic T cells specific for a LAGE-1 TAP comprising contacting an  
 CC isolated population of T cells with an agent presenting a complex of a  
 CC LAGE TAP or an immunogenic fragment and a HLA presenting molecule to  
 CC selectively enrich the isolated population of T cells with the cytolytic  
 CC T cells. The methods and products from the present invention can be used  
 CC for the diagnosis and treatment of LAGE-1 associated disorders,  
 CC particularly tumours.  
 XX  
 PS Sequence 210 AA;  
 SQ

Query Match 54.3%; Score 50; DB 19; Length 210;  
Best Local Similarity 54.5%; Pred. No. 31;  
Matches 12; Conservative 1; Mismatches 3; Indels 6; Gaps 2;

QY 1 PSHK-STQR-----PPPPGGRQ 16  
DB 171 PKHKVSEQRPGTGGPPPPGGAQ 192

RESULT 10  
AY70861  
ID AAY70861 standard; Protein; 210 AA.  
XX  
AC AAY70861;  
XX  
DT 31-JUL-2000 (first entry)  
XX  
DE Human LAGE-1 unspliced variant, LAGE-11 protein.  
XX  
LAGE-11; CAMEL; CTL-recognised Antigen on MELanoma; human; cancer; CTL;  
KW cytotoxic T lymphocyte; tumour-associated antigen; NY-ESO-1; anticancer;  
KW melanoma; immunotherapy; immune response.  
XX  
OS Homo sapiens.  
XX  
PN WO200023584-A1.  
XX  
PD 27-APR-2000.  
XX  
PF 15-OCT-1999; 99WO-EP07832.  
XX  
PR 16-OCT-1998; 98EP-0119583.  
XX  
PA (BOEH) BOEHRINGER INGELHEIM INT GMBH.  
PA (UYHO-) UNIV HOSPITAL LEIDEN.  
XX  
PI Schrier PI, Aarnoudse CA, Heider K, Klade C;  
XX  
WPI; 2000-339685/29.  
DR N-PSDB; AAD00151.  
XX  
Tumor-associated antigen useful for cancer immunotherapy is encoded by  
PT the open reading frame of LAGE-1 (a tumor-specific antigen) cDNA -  
PS Example 3; Page 60; 73pp; English.  
XX  
The present sequence is the human LAGE-1 unspliced variant, LAGE-11  
CC protein, a tumour-specific antigen. The LAGE-11 protein is encoded by the  
CC open reading frame-3 (ORF) and the CAMEL protein (Cytotoxic T lymphocytes  
CC (CTL)-recognised Antigen on MELanoma) from ORF-1 of LAGE-1 cDNA. The  
CC LAGE-1(s/l) protein derived from ORF-1 is identical to the CAMEL protein.  
CC It is not expressed in healthy tissues except in testis and placenta.  
CC It also shows homology with the NY-ESO-1, a tumour specific antigen. The  
CC tumour-associated antigen displayed on melanoma cells is recognised by  
CC cytotoxic T lymphocytes. This sequence has anticancer activity. CAMEL  
CC tumour antigen and immunogenic peptides derived from it are useful for  
CC cancer immunotherapy. They have the potential to induce an immune  
CC response, by eliciting a CTL response. The DNA molecule is used for the  
CC construction of recombinant or fusion proteins.  
XX  
SQ Sequence 210 AA;

Query Match 54.3%; Score 50; DB 21; Length 210;  
Best Local Similarity 54.5%; Pred. No. 31;  
Matches 12; Conservative 1; Mismatches 3; Indels 6; Gaps 2;

QY 1 PSHK-STQR-----PPPPGGRQ 16  
DB 171 PKHKVSEQRPGTGGPPPPGGAQ 192

RESULT 11  
AA08261

ID AA08261 standard; Protein; 114 AA.  
XX  
AC AA08261;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human polypeptide SEQ ID NO 22153.  
XX  
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorders; arthritis; inflammation.  
XX  
OS Homo sapiens.  
XX  
PN WO200164835-A2.  
XX  
PD 07-SEP-2001.  
XX  
PF 26-FEB-2001; 2001WO-US04927.  
XX  
PR 28-FEB-2000; 2000US-0515126.  
PR 18-MAY-2000; 2000US-0577409.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
Tang YT, Liu C, Drmanac RT;  
XX  
WPI; 2001-514838/56.  
DR N-PSDB; AAI88192.  
XX  
Isolated nucleic acids and polypeptides, useful for preventing  
PT diagnosing and treating e.g. leukaemia, inflammation and immune  
PT disorders -  
XX  
Claim 20; SEQ ID NO 22153; 1399pp + Sequence Listing; English.  
XX  
The invention relates to human polynucleotides (AAI79941-AAI93841) and  
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 114 AA;

Query Match 53.3%; Score 49; DB 22; Length 114;  
Best Local Similarity 53.3%; Pred. No. 24;  
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 PSHKSTQRPPPPGGR 15  
DB 70 PKNWTSPPPPPGGR 84

RESULT 12  
AAY48240  
ID AAY48240 standard; Protein; 167 AA.  
XX  
AC AAY48240;  
XX  
DT 08-DEC-1999 (first entry)  
XX  
DE Human prostate cancer-associated protein 26.  
XX

KW Expressed sequence tag; EST; prostate tumor; antitumor; treatment;  
 KW gene therapy; tissue specificity human.  
 XX Homo sapiens.  
 XX DE19811193-A1.  
 XX 16-SEP-1999.  
 XX 10-MAR-1998; 98DE-1011193.  
 XX 10-MAR-1998; 98DE-1011193.  
 XX (META-) METAGEN GES GENOMFORSCHUNG MBH.  
 XX Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E, Rosenthal A;  
 XX WPI; 1999-519628/44.  
 DR N-PSDB; AA233442.  
 XX New nucleic acid expressed at high level in prostatic tumor tissue and  
 PT encoded polypeptides, useful for treating cancer and screening for  
 PT therapeutic agents -  
 XX Claim 22; 124; 166pp; German.  
 PS This invention describes novel nucleic acid sequences (A) that are  
 CC expressed at high level in prostatic tumor tissue and encode gene  
 CC products or their fragments. The products of the invention have  
 CC antitumor activity. Polypeptides (i) encoded by (A) are used: (i) for  
 CC identifying agents for treatment of prostatic cancer and (ii) for  
 CC therapy of prostate cancer, optionally where expressed by gene therapy  
 CC methods. (A) is also used to isolate full-length genes (for gene therapy)  
 CC and for recombinant production of (i), which can be used to raise  
 CC specific antibodies. (A) are identified by assembly of ESTs (expressed  
 CC sequence tags) before they are analyzed for expression pattern (tissue  
 CC specificity). This approach eliminates many of the false results, as  
 CC regards tissue specificity, associated with known methods that use  
 CC single (usually short) ESTs. AA48215-49303 represent protein fragments  
 CC encoded by the expressed sequence tags described in the method of the  
 CC invention.  
 XX  
 SQ Sequence 167 AA;  
 Query Match 53.3%; Score 49; DB 20; Length 167;  
 Best Local Similarity 100.0%; Pred. No. 34;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 9 PPPPQGRQ 16  
 Db 10 PPPPQGRQ 17  
 RESULT 13  
 ABB68068  
 ID ABB68068 standard; Protein; 301 AA.  
 XX ABB68068;  
 AC ABB68068;  
 XX 26-MAR-2002 (first entry)  
 DT Drosophila melanogaster polypeptide SEQ ID NO 30996.  
 XX Drosophila melanogaster polypeptide; cell signalling; insecticide;  
 KW pharmaceutical.  
 XX Drosophila melanogaster.  
 XX WO200171042-A2.  
 PN 27-SEP-2001.  
 XX 23-MAR-2001; 2001WO-US09231.  
 PF

XX 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX (PEKE ) PE CORP NY.  
 XX Venter JC, Adams M, Li PWD, Myers EW;  
 PI WPI; 2001-656860/75.  
 DR N-PSDB; ABL12171.  
 XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX Disclosure; SEQ ID NO 30996; 21pp + Sequence Listing; English.  
 XX The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (ABB57737-ABB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 301 AA;  
 Query Match 53.3%; Score 49; DB 22; Length 301;  
 Best Local Similarity 66.7%; Pred. No. 59;  
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 PSKSKSTQRPPPP 12  
 Db 103 PADASTRRPPPP 114  
 RESULT 14  
 AAE23386  
 ID AAE23386 standard; Protein; 440 AA.  
 XX AAE23386;  
 AC AAE23386;  
 XX 27-AUG-2002 (first entry)  
 DT Human intracellular signalling (INTSIG-13) protein.  
 XX Human; intracellular signalling; INTSIG-13; gene therapy; nausea; asthma;  
 KW Crohn's disease; keratosis; cell proliferative disorder; Grave's disease;  
 KW arteriosclerosis; hepatitis; leukaemia; autoimmune disorder; meningitis;  
 KW acquired immunodeficiency syndrome; AIDS; allergy; dermatitis; psoriasis;  
 KW Hashimoto's thyroiditis; neurological disorder; developmental disorder;  
 KW gastrointestinal disorder; reproductive disorder; nootropic; anxiolytic;  
 KW Alzheimer's disease; dementia; Creutzfeldt-Jacob disease; dermatological;  
 KW cancer; neuroprotective; immunosuppressive; thyromimetic;  
 KW gynaecological.  
 XX Homo sapiens.  
 OS WO200231152-A2.  
 PN 18-APR-2002.  
 XX 12-OCT-2001; 2001WO-US32090.  
 PF 13-OCT-2000; 2000US-240871P.  
 XX 30-OCT-2000; 2000US-244723P.  
 PR 15-NOV-2000; 2000US-249402P.  
 XX 22-NOV-2000; 2000US-252622P.  
 PF

PR 13-DEC-2000; 2000US-255622P.  
XX (INCY-) INCYTE GENOMICS INC.  
PA  
XX Baughn MR, Ding L, Elliott VS, Gandhi AR, Gietzen KJ, Griffin JA;  
XX Gururajan R, Hafalia AUA, Kearney L, Khan FA, Lal P, Lee EA;  
PI Lu DAM, Lu Y, Nguyen DB, Arvizu C, Ramkumar J, Tang YT;  
PI Thangavelu K, Thornton M, Walia NK, Warren BA, Xu Y, Yao MG;  
PI Yue H;  
XX  
XX WPI; 2002-426285/45.  
DR N-PSDB; AAD37602.  
XX  
XX Novel intracellular signaling polypeptide useful for treating and  
XX preventing cell proliferative, autoimmune, neurological,  
XX gastrointestinal and reproductive and developmental disorders -  
XX  
XX Claim 68; Page 144-145; 177pp; English.  
XX  
XX The invention relates to an isolated intracellular signalling (INTSIG)  
XX polypeptide or a biologically active or immunogenic fragment of INTSIG.  
XX INTSIG is also useful for preparing a polyclonal or monoclonal antibody  
XX by hybridoma technology. INTSIG gene is useful in gene therapy. INTSIG  
XX is useful for treating or preventing disorders such as cell proliferative  
XX disorder e.g. keratosis, arteriosclerosis, hepatitis, psoriasis,  
XX leukaemia; autoimmune disorders e.g. AIDS, allergies, anaemia, asthma,  
XX Crohn's disease, dermatitis, Graves' disease, Hashimoto's thyroiditis;  
XX neurological disorders e.g. Alzheimer's disease, dementia, meningitis,  
XX Creutzfeldt-Jacob disease; gastrointestinal disorders e.g. nausea,  
XX reproductive disorders e.g. menstrual disorders, ectopic pregnancy and  
XX cancer of the breast and developmental disorders. The present sequence  
XX is human INTSIG-13 protein.  
XX  
XX Sequence 440 AA;  
SQ  
Query Match 53.3%; Score 49; DB 23; Length 440;  
Best Local Similarity 57.1%; Pred. No. 84;  
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
QY 3 HKSTORPPPPQGRQ 16  
| : |||||  
Db 172 HSSSAPPPPPGRR 185  
RESULT 15  
ABG01174  
ID ABG01174 standard; Protein; 442 AA.  
XX  
XX AC ABG01174;  
XX  
XX DT 13-FEB-2002 (first entry)  
XX  
XX DE Novel human diagnostic protein #1165.  
XX  
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
XX OS Homo sapiens.  
XX  
XX PN WO200175067-A2.  
XX  
XX PD 11-OCT-2001.  
XX  
XX PF 30-MAR-2001; 2001WO-US08631.  
XX  
XX PR 31-MAR-2000; 2000US-0540217.  
XX  
XX PR 23-AUG-2000; 2000US-0649167.  
XX  
XX PA (HYSE-) HYSEQ INC.  
XX  
XX PI Drmanac RT, Liu C, Tang YT;  
XX  
XX WPI; 2001-639362/73.  
XX

DR N-PSDB; AAS65361.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits and to assess  
XX biodiversity -  
XX  
XX Claim 20; SEQ ID No 31533; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,  
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX and gene mapping, and in recombinant production of (II). The  
XX polynucleotides are also used in diagnostics as expressed sequence tags  
XX for identifying expressed genes. (I) is useful in gene therapy techniques  
XX to restore normal activity of (II) or to treat disease states involving  
XX (II). (II) is useful for generating antibodies against it, detecting or  
XX quantitating a polypeptide in tissue, as molecular weight markers and as  
XX a food supplement. (II) and its binding partners are useful in medical  
XX imaging of sites expressing (II). (I) and (II) are useful for treating  
XX disorders involving aberrant protein expression or biological activity.  
XX The polypeptide and polynucleotide sequences have applications in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits to assess biodiversity  
XX and to produce other types of data and products dependent on DNA and  
XX amino acid sequences. ABG00010-ABG30377 represent novel human  
XX diagnostic amino acid sequences of the invention.  
XX Note: The sequence data for this patent did not appear in the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 442 AA;  
SQ  
Query Match 53.3%; Score 49; DB 22; Length 442;  
Best Local Similarity 66.7%; Pred. No. 84;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
QY 1 PSHKSTORPPPP 12  
| : |||||  
Db 27 PVHKTTSRDP 38  
| : |||||

Search completed: December 3, 2002, 14:28:32  
Job time : 35.6 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 3, 2002, 14:26:06 ; Search time 14.4 seconds  
(without alignments)  
106.816 Million cell updates/sec

Title: US-09-924-654-4\_COPY\_550\_565

Perfect score: 92

Sequence: 1 PSHKSTQRPQPQGRQ 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	52	56.5	629	2 E47096	membrane translocator
2	51	55.4	188	2 D29149	proline-rich prote
3	49.5	53.8	628	1 A56707	protein-tyrosine k
4	49	53.3	359	2 T13478	hypothetical prote
5	49	53.3	379	2 S31719	proline-rich prote
6	49	53.3	440	2 JC7807	Wiskott-Aldrich sy
7	49	53.3	708	2 D96711	hypothetical prote
8	49	53.3	963	2 T19140	hypothetical prote
9	48	52.2	1353	2 T00249	ichi protein - ink
10	47	51.1	240	2 B24264	proline-rich prote
11	47	51.1	564	2 T45866	hypothetical prote
12	47	51.1	600	2 S07838	spore coat protein
13	47	51.1	605	1 Q0B53R	BVRP2 (EC-RP3) pro
14	47	51.1	641	2 A45054	probable intercell
15	47	51.1	779	2 T20654	hypothetical prote
16	47	51.1	977	2 T16232	hypothetical prote
17	47	51.1	1072	2 T50949	verprolin related
18	47	51.1	1922	1 S02771	myosin heavy chain
19	47	51.1	2148	1 A56081	insulin receptor -
20	46.5	50.5	836	2 T08424	twooty protein - f
21	46	50.0	103	2 T29597	hypothetical prote
22	46	50.0	260	2 T23273	proline-rich prote
23	46	50.0	275	2 T02559	probable spliceoso
24	46	50.0	482	2 S65760	dihydrolipoamide t
25	46	50.0	506	1 S31229	probable RNA helic
26	46	50.0	764	2 H71607	hypothetical prote
27	46	50.0	786	2 T01456	extensin homolog F
28	46	50.0	871	2 T48502	hypothetical prote
29	46	50.0	1974	2 T30010	hypothetical prote

30	46	50.0	4135	2 T42629	tenascin-X - bovin
31	45	48.9	171	2 H70527	probable PE protei
32	45	48.9	227	2 C29149	proline-rich prote
33	45	48.9	233	2 AG2538	hypothetical prote
34	45	48.9	240	2 A24264	proline-rich prote
35	45	48.9	278	2 B39066	proline-rich prote
36	45	48.9	310	1 PIHUSD	salivary proline-r
37	45	48.9	317	2 A28996	proline-rich prote
38	45	48.9	366	2 T26449	hypothetical prote
39	45	48.9	393	2 PQ0479	pisil extensin-li
40	45	48.9	420	2 T42616	envelope protein -
41	45	48.9	503	2 C84595	similar to PEARLI
42	45	48.9	505	2 S72273	actin-depolymerizi
43	45	48.9	596	2 T03908	hypothetical prote
44	45	48.9	636	2 T38473	probable serine/th
45	45	48.9	681	2 F85062	hypothetical prote

ALIGNMENTS

RESULT 1

E47096 membrane translocator homolog amfB - Streptomyces griseus

C:Species: Streptomyces griseus

C:Date: 26-May-1994 #sequence\_revision 26-May-1994 #text\_change 02-Feb-2001

C:Accession: E47096

R:Ueda, K.; Miyake, K.; Horinouchi, S.; Beppu, T.

J. Bacteriol. 175, 2006-2016, 1993

S. A:Title: A gene cluster involved in aerial mycelium formation in Streptomyces griseus en

S. A:Reference number: A47096; MUID:93209944; PMID:8458843

A:Accession: E47096

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-629 <UED>

A:Cross-references: GB:D13614; NID:G3721543; PIDN:BAA33538.1; PID:G303911

C:Genetics:

A:Gene: amfB

A:Start codon: GTG

C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology

C:Keywords: ATP; nucleotide binding; P-loop

F:351-543/Domain: ATP-binding cassette homology <ABC>

F:368-375/Region: nucleotide-binding motif A (P-loop)

Query Match 56.5%; Score 52; DB 2; Length 629;

Best Local Similarity 50.0%; Pred. No. 13;

Matches 10; Conservative 3; Mismatches 3; Indels 4; Gaps 1;

QY 1 PSHKSTQRP-PPQGRQ 16

DB 604 PRHPTQRPQAPRPKSR 623

RESULT 2

D29149 proline-rich protein - mouse (fragment)

C:Species: Mus musculus (house mouse)

C:Date: 08-Mar-1989 #sequence\_revision 08-Mar-1989 #text\_change 20-Aug-1999

C:Accession: D29149

R:Clements, S.; Mehansho, H.; Carlson, D.M.

J. Biol. Chem. 260, 13471-13477, 1985

A:Title: Novel multigene families encoding highly repetitive peptide sequences. Sequence

A:Reference number: A92501; MUID:86033799; PMID:3840480

A:Contents: Clone pUM4

A:Accession: D29149

A:Molecule type: mRNA

A:Residues: 1-188 <CLE>

A:Cross-references: GB:M19419; NID:G200542; PIDN:AAA40002.1; PID:G200543

C:Superfamily: proline-rich protein

Query Match 55.4%; Score 51; DB 2; Length 188;

Best Local Similarity 56.2%; Pred. No. 5.4;

Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 PSHKSTQRPDPGQ 16  
 | : | | | | | | | |  
 Db 69 PQORSPQSPPPGPGQ 84

RESULT 3  
 A56707  
 protein-tyrosine kinase (EC 2.7.1.112) syk, splice form B - rat  
 N:Contains: protein-tyrosine kinase syk, splice form A  
 C:Species: Rattus norvegicus (Norway rat)  
 C>Date: 19-Oct-1995 #sequence\_revision 23-May-1997 #text\_change 10-Sep-1999  
 C:Accession: A56707; B56707; A48875  
 R:Rowley, R.B.; Bolen, J.B.; Fagnoli, J.  
 J. Biol. Chem. 270, 12659-12664, 1995  
 A:Title: Molecular cloning of rodent p72(Syk). Evidence of alternative mRNA splicing.  
 A:Reference number: A56707; MUID:95279402; PMID:7759516  
 A:Accession: A56707  
 A>Status: preliminary; not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1-628 <ROW>  
 A:Cross-references: GB:U01684  
 A:Accession: B56707  
 A>Status: preliminary; not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1-275,299-628 <RO2>  
 A:Cross-references: GB:U01683  
 R:Benhamou, M.; Ryba, N.J.P.; Nishikata, H.; Kihara, H.; Siraganian, R.P.  
 J. Biol. Chem. 268, 23318-23324, 1993  
 A:Title: Protein tyrosine kinase p72syk in high affinity IgE receptor signaling: identification  
 A:Reference number: A48875; MUID:94043123; PMID:7693687  
 A:Accession: A48875  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 15-22,'E',23-263,'M',265-269,'S',271-444,'EL',447-628 <RES>  
 A:Cross-references: GB:L20838; NID:9416152; PIDN:AAA42308.1; PID:9416153  
 C:Superfamily: protein-tyrosine kinase ZAP-70; protein kinase homology; SH2 homology  
 C:Keywords: alternative splicing; ATP; autophosphorylation; phosphoprotein; phosphotrans  
 F:1-628/Protein-tyrosine kinase syk, splice form B #status predicted <PR>  
 F:1-275,299-628/Protein-tyrosine kinase syk, splice form A #status predicted <P>  
 F:14-105/Domain: SH2 homology <SH2A>  
 F:166-257/Domain: SH2 homology <SH2B>  
 F:362-626/Domain: protein kinase homology <KIN>  
 F:370-378/Region: protein kinase ATP-binding motif

Query Match 53.8%; Score 49.5; DB 1; Length 628;  
 Best Local Similarity 64.3%; Pred. No. 28;  
 Matches 9; Conservative 1; Mismatches 1; Indels 3; Gaps 1;

QY 1 PSHKSTQRPDPGQ 14  
 | : | | | | | | | |  
 Db 294 PGHK---KPPPPGQ 304

RESULT 4  
 T13478  
 hypothetical protein 34F3.10 - fruit fly (Drosophila melanogaster)  
 C:Species: Drosophila melanogaster  
 C>Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 17-Nov-2000  
 C:Accession: T13478  
 R:Valenti, P.; Salles, C.; Campbell, L.; Glover, D.  
 submitted to the EMBL Data Library, April 1999  
 A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.  
 A:Reference number: Z17685  
 A:Accession: T13478  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-359 <PHI>  
 A:Cross-references: EMBL:AL031583; NID:el321005; PID:el321018; PIDN:CAB41346.1  
 C:Genetics:  
 A:Cross-references: FlyBase:FBgn0025623

A; Introns: 17/2; 50/3; 333/2  
 A; Note: EG:34F3.10

Query Match 53.3%; Score 49; DB 2; Length 359;  
 Best Local Similarity 66.7%; Pred. No. 19;  
 Matches 8; Conservative 2; Mismatches 0; Gaps 0;

QY 1 PSHKSTQRPDPGQ 12  
 | : | | | | | | | |  
 Db 161 PADASTRRPPPP 172

## RESULT 5

S31719  
 proline-rich protein - African clawed frog  
 C:Species: Xenopus laevis (African clawed frog)  
 C>Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 24-Nov-1999  
 C:Accession: S31719  
 R:Nishimatsu, S.I.; Satoshi, X.Y.Z.; Oda, X.Y.Z.; Naoto, X.Y.Z.; Ueno, X.Y.Z.  
 submitted to the EMBL Data Library, August 1992  
 A:Reference number: S31719  
 A:Accession: S31719  
 A>Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-379 <NIS>  
 A:Cross-references: EMBL:X68249; NID:964955; PID:964956  
 C:Superfamily: proline-rich protein

Query Match 53.3%; Score 49; DB 2; Length 379;  
 Best Local Similarity 57.1%; Pred. No. 20;  
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 HKSTQRPDPGQ 16  
 | : | | | | | | | |  
 Db 178 HSSSAPPDPGRR 191

## RESULT 6

JC7807  
 Wiskott-Aldrich syndrome protein (WASP)-interacting protein (WIP)-and CR16-homologous p;  
 C:Species: Homo sapiens (man)  
 C>Date: 03-Jun-2002 #sequence\_revision 03-Jun-2002 #text\_change 03-Jun-2002  
 C:Accession: JC7807  
 R:Kato, M.; Miki, H.; Kurita, S.; Endo, T.; Nakagawa, H.; Miyamoto, S.; Takenawa, T.  
 Biochem. Biophys. Res. Commun. 291, 41-47, 2002  
 A:Title: WICH, a novel verprolin homology domain-containing protein that functions coop  
 A:Reference number: JC7807; PMID:11829459; MUID:21818424  
 A:Contents: Jurkat T-cell  
 A:Accession: JC7807  
 A:Molecule type: mRNA  
 A:Residues: 1-440 <KAT>  
 C:Comment: This protein, a novel protein containing a verprolin-homology (V) domain, pl  
 e protein) family proteins.  
 C:Genetics:  
 A:Gene: wich

Query Match 53.3%; Score 49; DB 2; Length 440;  
 Best Local Similarity 57.1%; Pred. No. 23;  
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 HKSTQRPDPGQ 16  
 | : | | | | | | | |  
 Db 172 HSSSAPPDPGRR 185

## RESULT 7

D96711  
 hypothetical protein P24J5.8 [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
 C:Accession: D96711  
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso  
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Dewar, K.

ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408: 816-820, 2000  
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziala,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A;Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.  
A;Reference number: A86141; MUID:21016719; PMID:11130712  
A;Accession: D96711  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-708 <STO>  
A;Cross-references: GB:AE005173; NID:g5734709; PIDN:RAD49974.1; GSPDB:GN00141  
C;Genetics:  
A;Gene: F24J5.8  
A;Map position: 1

Query Match 53.3%; Score 49; DB 2; Length 708;  
Best Local Similarity 66.7%; Pred. No. 36;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
QY 1 PSHKSTORPPPP 12  
||:|||||  
Db 188 PSEPTQSPPPP 199

RESULT 8  
T19140  
hypothetical protein C09G5.6 - *Caenorhabditis elegans*  
C;Species: *Caenorhabditis elegans*  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C;Accession: T19140  
R;Palmer, S.  
A;Reference number: Z19080  
A;Submitted to the EMBL Data Library, November 1994  
A;Accession: T19140  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-963 <WIL>  
A;Cross-references: EMBL:Z46791; PIDN:CAA86755.1; GSPDB:GN00020; CESP:C09G5.6  
A;Experimental source: clone C09G5  
C;Genetics:  
A;Gene: CESP:C09G5.6  
A;Map position: 2  
A;Introns: 48/3; 862/3; 898/1

Query Match 53.3%; Score 49; DB 2; Length 963;  
Best Local Similarity 37.0%; Pred. No. 49;  
Matches 10; Conservative 2; Mismatches 3; Indels 12; Gaps 1;  
QY 1 PSHKST-----QRPPPPQGR 15  
||:|||||  
Db 665 PPHQTPELYPEQPVRRPPPPQNR 691

RESULT 9  
T00249  
ich1 protein - inky cap (*Coprinus cinereus*)  
C;Species: *Coprinus cinereus*  
C;Date: 22-Jan-1999 #sequence\_revision 22-Jan-1999 #text\_change 21-Jul-2000  
C;Accession: T00249  
R;Muraguchi, H.; Kanada, T.  
Development 125, 3133-3141, 1998  
A;Title: The ich1 gene of the mushroom *Coprinus cinereus* is essential for pileus formation  
A;Reference number: Z14131; MUID:98337832; PMID:9671586  
A;Accession: T00249  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-1353 <MUR>  
A;Cross-references: EMBL:AB012131; NID:g35551166; PIDN:BA432788.1; PID:g35551167  
A;Experimental source: strain 5302  
C;Genetics:

A;Gene: ich1  
A;Introns: 189/3; 233/1; 279/2; 395/2; 471/1; 527/3; 546/1; 663/3

Query Match 52.2%; Score 48; DB 2; Length 1353;  
Best Local Similarity 55.0%; Pred. No. 93;  
Matches 11; Conservative 1; Mismatches 4; Indels 4; Gaps 1;  
QY 1 PSHKSTORPP-----PPOGRO 16  
||:|||||  
Db 711 PGVSMRMRPPPREKRPPPOGRQ 730

RESULT 10  
B24264  
proline-rich protein MF3 - mouse (fragment)  
C;Species: *Mus musculus* (house mouse)  
C;Date: 09-Sep-1987 #sequence\_revision 09-Sep-1987 #text\_change 03-May-1996  
C;Accession: B24264  
R;Ann, D.K.; Carlson, D.M.  
J. Biol. Chem. 260, 15863-15872, 1985  
A;Title: The structure and organization of a proline-rich protein gene of a mouse multigene family.  
A;Reference number: A92508; MUID:86059475; PMID:2999141  
A;Accession: B24264  
A;Molecule type: DNA  
A;Residues: 1-240 <ANN>  
C;Superfamily: proline-rich protein

Query Match 51.1%; Score 47; DB 2; Length 240;  
Best Local Similarity 56.2%; Pred. No. 24;  
Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;  
QY 1 PSHKSTORPPPPPOGRQ 16  
||:|||||  
Db 135 PHLRPTQGPPOGPGQ 150

RESULT 11  
T45866  
hypothetical protein F3A4.210 - *Arabidopsis thaliana*  
C;Species: *Arabidopsis thaliana* (mouse-ear cress)  
C;Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 04-Feb-2000  
C;Accession: T45866  
R;Bargues, M.; Collado, M.C.; Navarro, P.; Terol, J.; Perez-Alonso, M.; Mewes, H.W.; May  
submitted to the Protein Sequence Database, December 1999  
A;Reference number: Z23007  
A;Accession: T45866  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-564 <BAR>  
A;Cross-references: EMBL:AL132978  
A;Experimental source: cultivar Columbia; BAC clone F3A4  
C;Genetics:  
A;Map position: 3  
A;Introns: 250/2; 424/1  
A;Note: F3A4.210

Query Match 51.1%; Score 47; DB 2; Length 564;  
Best Local Similarity 66.7%; Pred. No. 54;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
QY 1 PSHKSTORPPPP 12  
||:|||||  
Db 17 PPSRSIPRPPPP 28

RESULT 12  
S07638  
spore coat protein SP96 precursor - slime mold (*Dictyostelium discoideum*)  
C;Species: *Dictyostelium discoideum*  
C;Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 29-Oct-1999  
C;Accession: S07638; A60942; B60942  
R;Fosnaugh, K.L.; Loomis, W.F.  
Nucleic Acids Res. 17, 9489, 1989



A;Title: Sequence of the Dictyostelium discoideum spore coat gene SP96.  
A;Reference number: S07638; MUID:90067962; PMID:2587278  
A;Accession: S07638  
A;Molecule type: DNA  
A;Residues: 1-600 <FOS>  
A;Cross-references: EMBL:X16491; NID:g7373; PIDN:CAA34508.1; PID:g295736  
R;Tasaka, M.; Hasegawa, M.; Ozaki, T.; Iwabuchi, M.; Takeuchi, I.  
Cell Differ. Dev. 31, 1-9, 1990  
A;Title: Isolation and characterization of spore coat protein (sp96) gene of Dictyostelium  
A;Reference number: A60942; MUID:91028801; PMID:1977501  
A;Accession: A60942  
A;Molecule type: DNA  
A;Residues: 1-155, 'T', 157-414, 'C', 416-600 <TAS>  
A;Accession: B60942  
A;Status: not compared with conceptual translation  
A;Molecule type: mRNA  
A;Residues: 1-155, 'T', 157-414, 'C', 416-600 <TA2>  
C;Genetics:  
A;Introns: 22/1  
C;Keywords: duplication; glycoprotein  
F;1-18/Domain: signal sequence #status predicted <SIG>  
F;19-600/Product: spore coat protein SP96 #status predicted <MAT>  
F;142-185/Region: 15-residue repeats  
F;447-533/Region: 9-residue repeats  
F;568-600/Region: 32-residue repeats  
F;100/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 51.1%; Score 47; DB 2; Length 600;  
Best Local Similarity 58.3%; Pred. No. 57;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 PSHKSTQRPPOGRQ 12  
|:|:|:|:|:|:|  
Db 249 PTHRPTHKPP 260

RESULT 13  
Q08E3R  
BVRF2 (EC-RF3) protein - human herpesvirus 4 (strain B95-8)  
N;Contains: BVRF2 (EC-RF3a) protein  
C;Species: human herpesvirus 4, Epstein-Barr virus  
C;Date: 03-Aug-1984 #sequence\_revision 03-Aug-1984 #text\_change 16-Jul-1999  
C;Accession: A03798; S33049; S33050; S33051  
R;Bankier, A.T.; Deininger, P.L.; Farrell, P.J.; Barrell, B.G.  
Mol. Biol. Med. 1, 21-45, 1983  
A;Title: Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8 Epstein-Barr virus  
A;Reference number: A93065; MUID:85035713; PMID:6092825  
A;Accession: A03798  
A;Molecule type: DNA  
A;Residues: 1-605 <BAN>  
A;Cross-references: EMBL:V01555; NID:959074; PIDN:CAA24801.1; PID:G1334909  
R;Baer, R.; Bankier, A.T.; Biggin, M.D.; Deininger, P.L.; Farrell, P.J.; Gibson, T.J.; H  
Nature 310, 207-211, 1984  
A;Title: DNA sequence and expression of the B95-8 Epstein-Barr virus genome.  
A;Reference number: A03794; MUID:84270667; PMID:6087149  
A;Contents: annotation; protein coding region  
C;Superfamily: varicella-zoster virus gene 33 protein  
F;261-605/Product: BVRF2 (EC-RF3a) protein #status predicted <BVP>

Query Match 51.1%; Score 47; DB 1; Length 605;  
Best Local Similarity 53.3%; Pred. No. 58;  
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 2 SHKSTQRPPOGRQ 16  
|:|:|:|:|:|:|  
Db 508 SHQAAQRPPOPGTQ 522

RESULT 14  
A45034  
probable intercellular signal transducer or transmitter Fz-1 - rat  
C;Species: Rattus norvegicus (Norway rat)  
C;Date: 30-Apr-1993 #sequence\_revision 18-Nov-1994 #text\_change 02-Mar-2001

C;Accession: A45054  
R;Chan, S.D.; Karpf, D.B.; Fowlkes, M.E.; Hooks, M.; Bradley, M.S.; Vuong, V.; Bambino,  
J. Biol. Chem. 267, 25202-25207, 1992  
A;Title: Two homologs of the Drosophila polarity gene frizzled (fz) are widely expressed  
A;Reference number: A45054; MUID:93094228; PMID:1334084  
A;Accession: A45054  
A;Status: preliminary; not compared with conceptual translation  
A;Molecule type: nucleic acid  
A;Residues: 1-641 <CHA>  
A;Experimental source: UMR 106 osteosarcoma cell line  
A;Note: sequence extracted from NCBI backbone (NCBIP:120154)  
C;Superfamily: fruit fly frizzled protein

Query Match 51.1%; Score 47; DB 2; Length 641;  
Best Local Similarity 80.0%; Pred. No. 61;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 7 QRPPPOGRQ 16  
|:|:|:|:|:|:|  
Db 83 QRPPPOPGQ 92

RESULT 15  
T20654  
Hypothetical protein F09C8.2 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000  
C;Accession: T20654  
R;McMurray, A.  
Submitted to the EMBL Data Library, November 1995  
A;Reference number: Z19306  
A;Accession: T20654  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-779 <WIL>  
A;Cross-references: EMBL:Z68132; PIDN:CAA92220.1; GSPDB:GN00028; CESP:F09C8.2.  
C;Genetics:  
A;Gene: CESP:F09C8.2  
A;Map position: X  
A;Introns: 22/2; 165/1; 266/3; 306/3; 376/1; 462/1; 729/2

Query Match 51.1%; Score 47; DB 2; Length 779;  
Best Local Similarity 57.1%; Pred. No. 74;  
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 2 SHKSTQRPPOGRQ 15  
|:|:|:|:|:|:|  
Db 312 SHRSTKAPPPASR 325

Search completed: December 3, 2002, 14:29:07  
Job time : 16.4 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 3, 2002, 14:26:01 ; Search time 8 Seconds  
(without alignments)  
82.953 Million cell updates/sec

Title: US-09-924-654-4\_COPY\_550\_565

Perfect score: 92

Sequence: 1 PSHKSTQRPQPGRQ 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	52	56.5	1362	1 BRD4 HUMAN	O60885 homo sapien
2	50	54.3	210	1 CTG2 HUMAN	O75638 homo sapien
3	50	54.3	601	1 DR11 MOUSE	Q62431 mus musculus
4	49.5	53.8	629	1 KSYK RAT	Q64725 rattus norv
5	49	53.3	963	1 YQ36 CAEEL	Q09457 caenorhabdi
6	47	51.1	600	1 SP96 DICDI	P14328 dictyosteli
7	47	51.1	605	1 VP40 EBV	P03234 Epstein-bar
8	47	51.1	641	1 FZD1 RAT	Q08463 rattus norv
9	47	51.1	1969	1 MYSA CAEEL	P12844 caenorhabdi
10	46	50.0	482	1 ODB2 MOUSE	P53395 mus musculus
11	46	50.0	506	1 DH11 YEAST	P39517 saccharomyc
12	45.5	49.5	1051	1 YC94 HUMAN	O9p2q2 homo sapien
13	45	48.9	174	1 PRPP HUMAN	P81489 homo sapien
14	45	48.9	234	1 PRPM HUMAN	P10161 homo sapien
15	45	48.9	247	1 PRB4 HUMAN	P10163 homo sapien
16	45	48.9	261	1 PRP2 MOUSE	P05142 mus musculus
17	45	48.9	276	1 PRPL HUMAN	P10162 homo sapien
18	45	48.9	296	1 PRP3 MOUSE	P05143 mus musculus
19	45	48.9	501	1 WAS1 RAT	O08816 rattus norv
20	45	48.9	505	1 WAS1 BOVIN	Q95107 bos taurus
21	45	48.9	505	1 WASL HUMAN	O00401 homo sapien
22	45	48.9	636	1 KDGP SCHPO	O14019 schizosacch
23	45	48.9	707	1 SFPO HUMAN	P23246 homo sapien
24	45	48.9	1319	1 MN1 HUMAN	Q10571 homo sapien
25	45	48.9	2073	1 BIME EMENI	P24586 emericella
26	44.5	48.4	351	1 CD2 HUMAN	P06729 homo sapien
27	44.5	48.4	472	1 P2X2 RAT	P49653 rattus norv
28	44.5	48.4	1215	1 ATC6 YEAST	P39986 saccharomyc
29	44	47.8	166	1 PRPC HUMAN	P02810 homo sapien
30	44	47.8	443	1 OC3N HUMAN	P20265 homo sapien
31	44	47.8	443	1 OC3N MOUSE	P31360 mus musculus
32	44	47.8	445	1 OC3N RAT	P56222 rattus norv
33	44	47.8	662	1 NRG1 RAT	P43322 r pro-neure

RESULT 1

BRD4_HUMAN					
ID	BRD4_HUMAN	STANDARD;	PRT;	1362	AA.
AC	O60885	O96PD3;			
DT	16-OCT-2001	(Rel. 40, Created)			
DT	15-JUN-2002	(Rel. 41, Last sequence update)			
DT	15-JUN-2002	(Rel. 41, Last annotation update)			
DE	Bromodomain-containing protein 4 (HUNK1 protein).				
GN	BRD4 OR HUNK1.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	French C.A., Fletcher J.A.;				
RT	Human BRD4 protein.				
RL	Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.				
RN	[2]				
RP	SEQUENCE OF 1-722 FROM N.A.				
RC	TISSUE=Placenta;				
RA	Weber B.;				
RL	Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.				
CC	SUBCELLULAR LOCATION: Nuclear (Potential).				
CC	SIMILARITY: CONTAINS 2 BROMODOMAINS.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).				
CC	EMBL; AF386649; AAL26987.1; -				
CC	EMBL; Y12059; CAA72780.1; -				
DR	HSSP; Q92831; 1B91.				
DR	Genew; HGNC:13575; BRD4.				
DR	InterPro; IPR001487; Bromodomain.				
DR	Pfam; PF00439; bromodomain; 2.				
DR	PRINTS; PR00503; BROMODOMAIN.				
DR	SMART; SM00297; BROMO; 2.				
DR	PROSITE; PS00633; BROMODOMAIN_1; 1.				
DR	PROSITE; PS50014; BROMODOMAIN_2; 2.				
KW	Bromodomain; Repeat; Nuclear protein.				
FT	DOMAIN	75	147	BROMODOMAIN 1.	
FT	DOMAIN	368	440	BROMODOMAIN 2.	
FT	DOMAIN	535	594	LVS-RICH.	
FT	DOMAIN	692	717	SER-RICH.	
FT	DOMAIN	703	714	POLY-SER.	
FT	DOMAIN	738	743	POLY-HIS.	
FT	DOMAIN	757	761	POLY-PRO.	
FT	DOMAIN	764	770	POLY-PRO.	
FT	DOMAIN	771	775	POLY-GLN.	
FT	DOMAIN	776	783	POLY-PRO.	
FT	DOMAIN	954	964	POLY-PRO.	

P16065 strongyloce  
P27284 eastern equ  
Q92349 schizosacch  
O18964 bos taurus  
Q62910 rattus norv  
O63426 homo sapien  
P02567 caenorhabdi  
P12845 caenorhabdi  
P02566 caenorhabdi  
Q9ub19 homo sapien  
O61572 mus musculu  
O15027 homo sapien



```
Best Local Similarity 57.1%; Pred. No. 12;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 PSHKSTQRPDPG 14
||| :| |||
Db 213 PSHMASQMPDPHG 226

RESULT 4
KSYK RAT
ID KSYK RAT STANDARD; PRT; 629 AA.
AC Q64725;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase SYK (EC 2.7.1.12) (Spleen tyrosine kinase).
GN SYK.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95279402; PubMed=7759516;
RA Rowley R.B., Bolen J.B., Fargnoli J.;
RT "Molecular cloning of rodent p72Syk. Evidence of alternative mRNA
splicing.";
RL J. Biol. Chem. 270:12659-12664 (1995).
CC -!- FUNCTION: MAY PARTICIPATE IN SIGNALING PATHWAYS. PLAYS A ROLE IN
LYMPHOCYTE ACTIVATION.
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
tyrosine phosphate.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; SYK AND SYKB (SHOWN HERE);
CC ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- PTM: AUTOPHOSPHORYLATED (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE TYR FAMILY OF PROTEIN KINASES. SYK/ZAP-
70 SUBFAMILY.
CC -!- SIMILARITY: CONTAINS 2 SH2 DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; U21684; AAA75167.1; -
CC EMBL; U21683; AAA75166.1; -
CC HSP; P43405; IAB1.
CC InterPro; IPR000719; Euk_pkinase.
CC InterPro; IPR000980; SH2.
CC InterPro; IPR001245; Tyr_pkinase.
CC Pfam; PF00017; SH2; 2.
CC Pfam; PF00069; pkinase; 1.
CC PRINTS; PR00401; SH2DOMAIN.
CC ProDom; PD000001; Euk_pkinase; 1.
CC ProDom; PD000093; SH2; 2.
CC SMART; SM00252; SH2; 2.
CC SMART; SM00219; TyrKc; 1.
CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
CC PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
CC PROSITE; PS50001; SH2; 2.
CC Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW SH2 domain; Repeat; Alternative splicing.
FT DOMAIN 14 106
FT DOMAIN 167 258
FT DOMAIN 365 625
FT NP_BIND 371 379
FT BINDING 396 396
FT ACT_SITE 488 488
FT MOD_RES 519 519
PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
```

```
FT VARSPLIT 277 299 MISSING (IN ISOFORM SYKA).
SQ SEQUENCE 629 AA; 71528 MW; 81169A643EC6A6FE CRC64;

Query Match 53.8%; Score 49.5; DB 1; Length 629;
Best Local Similarity 64.3%; Pred. No. 15;
Matches 9; Conservative 1; Mismatches 1; Indels 3; Gaps 1;

QY 1 PSHKSTQRPDPG 14
||| :| |||
Db 295 PGHK---KPPPPQG 305

RESULT 5
YQ36 CAEEL
ID YQ36 CAEEL STANDARD; PRT; 963 AA.
AC Q09457;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative cuticle collagen C09G5.6.
GN C09G5.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Palmer S.;
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: NEMATODE CUTICLES ARE COMPOSED LARGELY OF COLLAGEN-LIKE
PROTEINS. THE CUTICLE FUNCTIONS BOTH AS AN EXOSKELETON AND AS A
BARRIER TO PROTECT THE WORM FROM ITS ENVIRONMENT (BY SIMILARITY).
CC -!- SUBUNIT: COLLAGEN POLYPEPTIDE CHAINS ARE COMPLEXED WITHIN THE
CUTICLE BY DISULFIDE BONDS AND OTHER TYPES OF COVALENT CROSS-
LINKS (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE CUTICULAR COLLAGEN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; Z46791; CAA86755.1; -
CC WormPep; C09G5.6; CE01486.
CC InterPro; IPR002486; Col_cuticle_N.
CC InterPro; IPR000087; Collagen.
CC Pfam; PF01391; Collagen; 3.
CC Pfam; PF01484; Col_cuticle_N; 1.
CC Hypothetical protein; Cuticle; Connective tissue; Repeat;
KW Multigene family; Cuticle.
FT DOMAIN 392 423
FT DOMAIN 441 503
FT DOMAIN 506 567
FT DOMAIN 663 666
FT DOMAIN 685 688
FT SEQUENCE 963 AA; 107031 MW; AFF895A75909F66E CRC64;

Query Match 53.3%; Score 49; DB 1; Length 963;
Best Local Similarity 57.0%; Pred. No. 26;
Matches 10; Conservative 2; Mismatches 3; Indels 12; Gaps 1;

QY 1 PSHKST-----QRPPPPQGR 15
||| :| |||
Db 665 PPHRTPHLYEPEQFYVRRPPPPQNR 691

RESULT 6
SP96_DICDI
ID SP96_DICDI STANDARD; PRT; 600 AA.
```

AC P14328;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE Spore coat protein SP96.  
GN COTA.  
OS Dictyostelium discoideum (Slime mold).  
OC Eukaryota; Mycotozoa; Dictyostelida; Dictyostelium.  
OX NCBI\_TaxID=44689;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=AX4;  
RX MEDLINE=90067962; PubMed=2587278;  
RA Fosnaugh K., Loomis W.F.;  
RT "Sequence of the Dictyostelium discoideum spore coat gene SP96."  
RL Nucleic Acids Res. 17:9489-9489(1989).  
CC -1- SUBCELLULAR LOCATION: OUTER LAYER OF THE COAT MATRIX AND  
CC INTERSPORE MATRIX.  
CC -1- PTM: PHOSPHORYLATED AND FUCOSYLATED. MAY BE PHOSPHOGLYCOSYLATED,  
CC MAY CONTAIN GLCNAC-ALPHA-1-P-SER RESIDUES.  
CC -1- SIMILARITY: CONTAINS 4 PRESPORE MOTIFS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL; X16491; CAA34508.1; -;  
DR PIR; S07638; S07638.  
DR GlycoSuiteDB; P14328; -;  
DR DictyDB; DD03007; COTA.  
DR InterPro; IPR003645; FOLN.  
DR SMART; SM00274; FOLN; 6  
KW Glycoprotein; Phosphorylation; Repeat; Sporulation.  
FT REPEAT 185 197 PRESPORE MOTIF 1.  
FT REPEAT 221 233 PRESPORE MOTIF 2.  
FT REPEAT 298 310 PRESPORE MOTIF 3.  
FT REPEAT 395 407 PRESPORE MOTIF 4.  
FT SEQUENCE 600 AA; 59589 MW; 616AE6D02B5F1071 CRC64;  
SQ  
Query Match 51.1%; Score 47; DB 1; Length 600;  
Best Local Similarity 58.3%; Pred. No. 30;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
QY 1 PSHKSTORPPPP 12  
DB 249 PTHRPTHRPKPP 260.  
  
RESULT 7  
ID VP40\_EBV STANDARD; PRT; 605 AA.  
AC P03234;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Capsid protein P40 (Virion structural protein BVRF2) (EC-RF3 and EC-  
DE RF3A) [Contains: Capsid protein VP24 (Assemblin) (Protease) (EC  
DE 3.4.21.97); Capsid protein VP22A].  
GN BVRF2.  
OS Epstein-Barr virus (strain B95-8) (Human herpesvirus 4).  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Gammapherpesvirinae; Lymphocryptovirus.  
OX NCBI\_TaxID=10377;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=85035713; PubMed=6092825;  
RA Bankier A.T., Deininger P.L., Farrell P.J., Barrell B.G.;  
RT "Sequence analysis of the 17,166 base-pair EcoRI fragment C of B95-8  
RT Epstein-Barr virus."

RL Mol. Biol. Med. 1:21-45(1983).  
CC -1- FUNCTION: VP22A IS A COMPONENT OF THE CAPSID CORE INVOLVED IN  
CC PROCESSING AND PACKAGING OF PROGENY DNA. VP24 IS A PROTEASE  
CC WHICH CAN PROTEOLYTICALLY CLEAVE ITSELF AND VP22A AT THE C-  
CC TERMINUS (BY SIMILARITY).  
CC -1- CATALYTIC ACTIVITY: Cleaves -Ala-|-Ser- and -Ala-|-Ala-bonds in  
CC the scaffold protein.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S21.  
CC -1- SIMILARITY: TO OTHER HERPESVIRUSES CAPSID PROTEIN VP40.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL; V01555; CAA24801.1; -;  
DR PIR; A03798; Q0BE3R.  
DR PIR; S33049; S33049.  
DR HSSP; P16753; 1WFO.  
DR MEROPS; S21.003; -;  
DR InterPro; IPR001847; Assemblin.  
DR Pfam; PF00716; Peptidase\_S21; 1.  
DR PRINTS; PR00236; HSCVAPSIDP40.  
KW Coat protein; Hydrolase; Serine protease; Alternative initiation.  
FT CHAIN 1 605 PROTEIN EC-RF3.  
FT CHAIN 261 605 PROTEIN EC-RF3A.  
FT CHAIN 1 235 COAT PROTEIN VP24 (PROTEASE).  
FT CHAIN 236 7568 COAT PROTEIN VP22A.  
FT PROPEP 7569 605 C-TERMINAL PEPTIDE.  
FT INIT MET 261 261 FOR PROTEIN EC-RF3A.  
FT SITE 235 236 CLEAVAGE (BY THE PROTEASE) (PROBABLE).  
FT SITE 568 569 CLEAVAGE (BY THE PROTEASE) (PROBABLE).  
FT ACT SITE 48 48 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT SITE 116 116 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT SITE 139 139 CHARGE RELAY SYSTEM (BY SIMILARITY).  
SQ SEQUENCE 605 AA; 64101 MW; FC2D355F9A389708 CRC64;  
Query Match 51.1%; Score 47; DB 1; Length 605;  
Best Local Similarity 53.3%; Pred. No. 30;  
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;  
QY 2 SHKSTORPPPPQGRQ 16  
DB 508 SHQAQAQPPPPGTQ 522  
  
RESULT 8  
ID FZD1\_RAT STANDARD; PRT; 641 AA.  
AC Q08463;  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Frizzled 1 precursor (Frizzled-1) (Fz-1) (rFz1).  
GN FZD1.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Sprague-Dawley; TISSUE=Osteosarcoma;  
RX MEDLINE=93094228; PubMed=1334084;  
RA Chan S.D.H., Karpf D.B., Fowlkes M.E., Hooks M., Bradley M.S.,  
RA Vuong V., Bambino T., Liu M.Y.C., Arnaud C.D., Strewler G.J.,  
RA Nissen R.A.;  
RT "Two homologs of the Drosophila polarity gene frizzled (fz) are widely  
RT expressed in mammalian tissues."  
RL J. Biol. Chem. 267:25202-25207(1992).  
RN [2]

RP COUPLING TO BETA-CATENIN PATHWAY.  
 RA MEDLINE=9324245; PubMed=10395542;  
 RX Sheidahl L.C., Park M., Malbon C.C., Moon R.T.;  
 RT "Protein kinase C is differentially stimulated by Wnt and Frizzled  
 homologs in a G-protein-dependent manner";  
 Curr. Biol. 9:695-698(1999).  
 RL -!- FUNCTION: Receptor for Wnt proteins. Most of frizzled receptors  
 are coupled to the beta-catenin canonical signaling pathway, which  
 leads to the activation of dishevelled proteins, inhibition of  
 GSK-3 kinase, nuclear accumulation of beta-catenin and activation  
 of Wnt target genes. A second signaling pathway involving PKC and  
 calcium fluxes has been seen for some family members, but it is  
 not yet clear if it represents a distinct pathway or if it can be  
 integrated in the canonical pathway, as PKC seems to be required  
 for Wnt-mediated inactivation of GSK-3 kinase. Both pathways seem  
 to involve interactions with G-proteins. May be involved in  
 transduction and intercellular transmission of polarity  
 information during tissue morphogenesis and/or in differentiated  
 tissues. Activation by Wnt8 induces expression of beta-catenin  
 target genes.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -!- TISSUE SPECIFICITY: Widely expressed. Most abundant in kidney,  
 liver, uterus, ovary and heart. Lower levels seen in brain and  
 intestine. Extremely low in calvaria, mammary glands and testis.  
 CC -!- DEVELOPMENTAL STAGE: Expressed predominantly in neonatal tissues,  
 at lower levels in adult.  
 CC -!- DOMAIN: Lys-Thr-X-X-X-Trip motif is involved in the activation of  
 the Wnt/beta-catenin signaling pathway (By similarity).  
 CC -!- DOMAIN: The fz domain is involved in binding with Wnt ligands (By  
 similarity).  
 CC -!- SIMILARITY: BELONGS TO FAMILY FZ/SMO OF G-PROTEIN COUPLED  
 RECEPTORS.  
 CC -!- SIMILARITY: CONTAINS 1 FRIZZLED (FZ) DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; L025229; AAA41173.1; -;  
 DR InterPro; IPR000539; Frizzled.  
 DR InterPro; IPR000024; Fz domain.  
 DR InterPro; IPR000832; GPCR\_secretin.  
 DR Pfam; PF01392; Fz; 1.  
 DR Pfam; PF01534; Frizzled; 1.  
 DR PRINTS; PR00489; FRIZZLED.  
 DR SMART; SM00063; FRI; 1.  
 DR PROSITE; PS50038; FZ; 1.  
 DR PROSITE; PS50261; G-PROTEIN RECEPTOR F2.4; 1.  
 KW Multigene family; G-protein coupled receptor; Transmembrane;  
 KW Developmental protein; Glycoprotein; Signal.  
 FT SIGNAL 1 68  
 FT CHAIN 69 641  
 FT DOMAIN 69 316  
 FT DOMAIN 317 337  
 FT DOMAIN 338 348  
 FT DOMAIN 349 369  
 FT DOMAIN 370 396  
 FT DOMAIN 397 417  
 FT DOMAIN 418 439  
 FT DOMAIN 440 460  
 FT DOMAIN 461 483  
 FT DOMAIN 484 504  
 FT DOMAIN 505 530  
 FT DOMAIN 531 551  
 FT DOMAIN 552 595  
 FT DOMAIN 596 616  
 FT DOMAIN 617 641  
 FT DOMAIN 106 224  
 FT DOMAIN 85 90  
 FT POLY-PRO.

FT SITE 619 624 LYS-THR-X-X-TRP MOTIF.  
 FT SITE 639 641 PDZ-BINDING.  
 FT CARBOHYD 125 125 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 225 225 N-LINKED (GLCNAC...) (POTENTIAL).  
 SQ SEQUENCE 641 AA; D82E2C113E81B8B6 CRC64;  
 Query Match 51.1%; Score 47; DB 1; Length 641;  
 Best Local Similarity 80.0%; Pred. No. 32;  
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 Qy 7 QRRPPPPQGRQ 16  
 Db 83 QRRPPPPQGRQ 92  
 RESULT 9  
 MYSA CAEEL STANDARD; PRT; 1969 AA.  
 AC P12844;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 01-OCT-1989 (Rel. 12, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Myosin heavy chain A (MHC A).  
 GN MYO-3.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN=Bristol N2;  
 RX MEDLINE=89178677; PubMed=2926820;  
 RA Dibb N.J., Maruyama I.N., Krause M., Karn J.;  
 RT "Sequence analysis of the complete Caenorhabditis elegans myosin  
 heavy chain gene family";  
 RL J. Mol. Biol. 205:503-513(1989).  
 CC -!- FUNCTION: MUSCLE CONTRACTION.  
 CC -!- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2  
 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)  
 AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).  
 CC -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.  
 CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING  
 CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,  
 CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.  
 CC -!- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY  
 ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.  
 CC -!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT  
 MEROMYOSIN (LMW) AND 1 HEAVY MEROMYOSIN (HWM). IT CAN LATER BE  
 SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED  
 SUBFRAGMENT (S2).  
 CC -!- MISCELLANEOUS: THERE ARE FOUR DIFFERENT MYOSIN HEAVY CHAINS IN  
 C. ELEGANS.  
 CC -!- MISCELLANEOUS: MHC A AND MHC B ARE FOUND EXCLUSIVELY IN THE BODY  
 WALL MUSCLE. THEY CO-ASSEMBLE INTO BODY WALL THICK FILAMENT.  
 CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.  
 CC -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; X08067; CAA30856.1; -;  
 DR PIR; S02771; S02771.  
 DR HSP; P08799; 1MND.  
 DR InterPro; IPR000048; IQ region.  
 DR InterPro; IPR004009; Myosin\_N.  
 DR InterPro; IPR002928; Myosin\_tail.  
 DR InterPro; IPR001609; myosin\_head.  
 DR Pfam; PF00063; myosin\_head; 1.

```
DR Pfam; PF00612; IQ; 1.
DR Pfam; PF01576; Myosin_tail; 1.
DR Pfam; PF02736; Myosin_N; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; IQSc; 1.
DR PROSITE; PS00096; IQ; 1.
KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW ATP-binding; Methylation; Alkylation; Multigene family.
FT DOMAIN 1 793
FT DOMAIN 794 823
FT DOMAIN 857 1969
FT NP_BIND 179 186
FT DOMAIN 667 689
FT DOMAIN 770 784
FT MOD_RES 130 130
FT MOD_RES 707 707
FT MOD_RES 717 717
SQ SEQUENCE 1969 AA; 225509 MW; 64577BBAF7EAD80A CRC64;

Query Match 51.1%; Score 47; DB 1; Length 1969;
Best Local Similarity 50.0%; Pred. No. 99;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 3 HKSTQRPQPQGRQ 16
DB 564 HPNFQPKPKGKQ 577

RESULT 10
ID_OB2_MOUSE STANDARD; PRT; 482 AA.
AC P53395;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 30-MAY-2000 (Rel. 39, Last annotation update)
DE Lipamide acyltransferase component of branched-chain alpha-keto acid
DE dehydrogenase complex, mitochondrial precursor (EC 2.3.1.-) (E2)
DE (Dihydrolipoamide branched chain transacylase) (BCKAD E2 subunit).
GN DBT.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=96180980; PubMed=8605244;
RA Costas P.A., Tonelli L.A., Chinsky J.M.;
RT "Molecular cloning of the murine branched chain alpha-ketoacid
RT dehydrogenase E2 subunit: presence of 3' E1 repeat elements.";
RL Biochim. Biophys. Acta 1305:25-28(1996).
CC -1- FUNCTION: THE BRANCHED-CHAIN ALPHA-KETO DEHYDROGENASE COMPLEX
CC CATALYZES THE OVERALL CONVERSION OF ALPHA-KETO ACIDS TO ACYL-COA
CC AND CO(2). IT CONTAINS MULTIPLE COPIES OF 3 ENZYMIC COMPONENTS:
CC BRANCHED-CHAIN ALPHA-KETO ACID DECARBOXYLASE (E1), LIPOAMIDE
CC ACYLTRANSFERASE (E2) AND LIPOAMIDE DEHYDROGENASE (E3).
CC -1- COFACTOR: THE E2 COMPONENT CONTAINS ONE COVALENTLY-BOUND LIPOYL
CC -1- SUBUNIT. FORMS A 24-POLYPEPTIDE STRUCTURAL CORE WITH OCTAHEDRAL
CC SYMMETRY.
CC -1- SUBCELLULAR LOCATION: Mitochondrial matrix.
CC -1- MISCELLANEOUS: THE CATALYTIC FUNCTION OF THIS ENZYME IS TO ACCEPT,
CC AND TO TRANSFER TO COENZYME A, ACYL GROUPS THAT ARE GENERATED BY
CC THE BRANCHED-CHAIN ALPHA-KETO ACID DECARBOXYLASE COMPONENT.
CC -1- SIMILARITY: BELONGS TO THE 2-OXOACID DEHYDROGENASE FAMILY.
CC -1- SIMILARITY: CONTAINS 1 LIPOYL-BINDING DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC CC
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC DR EMBL; L42996; AAC37681.1; -.
CC DR HSSP; P11961; 2PDD.
CC DR MGD; MGI:105386; Dbt.
CC InterPro; IPR001078; 2Oxoacid dh.
CC InterPro; IPR000089; Biotin lipoyl.
CC InterPro; IPR004167; E3 binding.
CC InterPro; IPR003016; Lipoyl.
CC Pfam; PF00198; 2-oxoacid dh; 1.
CC Pfam; PF0364; biotin lipoyl; 1.
CC Pfam; PF02817; e3 binding; 1.
CC ProDom; PD001115; 2Oxoacid dh; 1.
CC PROSITE; PS00189; LIPOYL_I.
KW Transferase; Acyltransferase; Mitochondrion; Transit peptide; Lipoyl.
FT TRANSIT 1 58
FT CHAIN 59 482
FT CHAIN 59 482
FT BINDING 105 105
FT ACT_SITE 452 452
FT ACT_SITE 456 456
FT ACT_SITE 456 456
SQ SEQUENCE 482 AA; 53160 MW; DBA87831B4BD2B01 CRC64;

Query Match 50.0%; Score 46; DB 1; Length 482;
Best Local Similarity 49.2%; Pred. No. 32;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 PSHKSTQRPQPQ 13
DB 219 PPKSEITPPQ 231

RESULT 11
ID_DHH1_YEAST STANDARD; PRT; 506 AA.
AC P39517;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Putative ATP-dependent RNA helicase DHH1.
GN DHH1 OR YDL160C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93289822; PubMed=8511971;
RA Strahl-Bolsinger S., Tanner W.;
RT "A yeast gene encoding a putative RNA helicase of the 'DEAD'-box
RT family.";
RL Yeast 9:429-432(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c;
RA Fohl T.M.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE DEAD BOX HELICASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC CC
CC EMBL; X66057; CAA46853.1; -.
CC EMBL; Z67750; CAA91586.1; -.
CC EMBL; Z74208; CAA98734.1; -.
-----
```

DR PIR; S31229; S31229.  
DR HSP; O58083; 1HV8.  
DR SGD; S0002319; DH1.  
DR InterPro; IPR001410; DEAD.  
DR InterPro; IPR000629; DEAD\_box.  
DR InterPro; IPR001650; Helicase\_C.  
DR Pfam; PF00270; DEAD; 1.  
DR Pfam; PF00271; Helicase\_C; 1.  
DR SMART; SM00487; DEXDC; 1.  
DR SMART; SM00490; HELIC; 1.  
DR PROSITE; PS00039; DEAD\_ATP\_HELICASE; 1.  
KW ATP-binding; RNA-binding; Helicase.  
FT NP BIND 90 97  
FT SITE 195 198 DEAD\_BOX.  
SQ SEQUENCE 506 AA; 57544 MW; 062CFAS6DF6F2CEE CRC64;  
Query Match 50.0%; Score 46; DB 1; Length 506;  
Best Local Similarity 61.5%; Pred. No. 34;  
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
QY 1 PSHKSTQRPDPQ 13  
DB 468 PSHQOQAYPPQ 480  
RESULT 12  
YC94 HUMAN  
ID YC94 HUMAN STANDARD; PRT; 1051 AA.  
AC Q9P2Q2;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical protein KIAA1294 (fragment).  
GN KIAA1294.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=20181126; PubMed=10718198;  
RA Nagase T., Kikuno R., Iehikawa K.-I., Hirosewa M., Ohara O.;  
RT "Prediction of the coding sequences of unidentified human genes XVI.  
RT The complete sequences of 150 new cDNA clones from brain which code  
RT for large proteins in vitro.";  
RL DNA Res. 7:65-73(2000).  
CC -!- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AB037715; BAA92532.1; -  
DR InterPro; IPR000299; Band 4.1.  
DR Pfam; PF00373; Band 41; 1.  
DR PRINTS; PR00935; BANO41.  
DR SMART; SM00295; B41; 1.  
DR PROSITE; PS00660; BAND\_41\_1; 1.  
DR PROSITE; PS00661; BAND\_41\_2; FALSE\_NEG.  
DR PROSITE; PS00557; BAND\_41\_3; 1.  
KW Hypothetical protein; Cytoskeleton.  
FT NON\_TER 1  
FT DOMAIN 29 253  
SQ SEQUENCE 1051 AA; 116890 MW; D1B16058C225E385 CRC64;  
Query Match 49.5%; Score 45.5; DB 1; Length 1051;  
Best Local Similarity 69.2%; Pred. No. 83;  
Matches 9; Conservative 0; Mismatches 1; Indels 3; Gaps 1;

QY 1 PSHKSTQRPDPQ 13  
DB 589 PSH---NRPPPPQ 598  
RESULT 13  
PRPP HUMAN  
ID PRPP HUMAN STANDARD; PRT; 174 AA.  
AC P81489;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE Salivary proline-rich protein II-1 (fragment).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE.  
RC TISSUE=Saliva;  
RX MEDLINE=93385383; PubMed=8373986;  
RA Kauffman D.L., Keller P.J., Bennick A., Blum M.;  
RT "Alignment of amino acid and DNA sequences of human proline-rich  
RT proteins.";  
RL Crit. Rev. Oral Biol. Med. 4:287-292(1993).  
KW Repeat; Parotid gland; Saliva; Multigene family.  
FT NON\_TER 174 174  
SQ SEQUENCE 174 AA; 17802 MW; D645F106EB1BB5BE CRC64;  
Query Match 48.9%; Score 45; DB 1; Length 174;  
Best Local Similarity 77.8%; Pred. No. 16;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 7 QRPDPDPQGR 15  
DB 31 QRPDPDPQGR 39  
RESULT 14  
PRPM HUMAN  
ID PRPM HUMAN STANDARD; PRT; 234 AA.  
AC P0161; P02813;  
DT 01-MAR-1989 (Rel. 10, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Salivary proline-rich protein PO (Allele M) [Contains: Peptide P-D]  
DE (Fragment).  
GN PRB4.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89121439; PubMed=3220251;  
RA Lyons K.M., Stein J.H., Smithies O.;  
RT "Many protein products from a few loci: assignment of human salivary  
RT proline-rich proteins to specific loci.";  
RL Genetics 120:255-265(1988).  
RN [2]  
RP SEQUENCE OF 165-234.  
RX MEDLINE=83186122; PubMed=6841349;  
RA Saitoh E., Isemura S., Sanada K.;  
RT "Complete amino acid sequence of a basic proline-rich peptide, P-D,  
RT from human parotid saliva.";  
RL J. Biochem. 93:495-502(1983).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial



CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
DR EMBL; X07704; CAA30542.1; -  
DR FIR; A03295; PIHUSD.  
DR PIR; S03175; S03175.  
DR MIM; 168730; -  
DR MIM; 180990; -  
KW Repeat; Parotid gland; Saliva; Multigene family.  
FT NON\_TER 1  
FT CHAIN 165 234 PEPTIDE P-D.  
SQ SEQUENCE 234 AA; 23676 MW; 310AFF13A44E747F CRC64;

Query Match 48.9%; Score 45; DB 1; Length 234;  
Best Local Similarity 77.8%; Pred. No. 21;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 7 QRPDPQGR 15  
Db 14 QRPDPDPGK 22

## RESULT 15

PRB4\_HUMAN STANDARD; PRT; 247 AA.  
ID PRB4\_HUMAN P02813;  
AC P10163; P02813;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 13-AUG-1987 (Rel. 05, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Salivary proline-rich protein PO precursor (Allele S).  
GN PRB4.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=85289325; PubMed=2993301;  
RA Maeda N., Kim H.-S., Azen E.A., Smithies O.;  
RT "Differential RNA splicing and post-translational cleavages in the human salivary proline-rich protein gene system.";  
RL J. Biol. Chem. 260:11123-11130(1985).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89121440; PubMed=2851479;  
RA Lyons K.M., Stein J.H., Smithies O.;  
RT "Length polymorphisms in human proline-rich protein genes generated by intragenic unequal crossing over.";  
RL Genetics 120:267-278(1988).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

Query Match 48.9%; Score 45; DB 1; Length 247;

Best Local Similarity 77.8%; Pred. No. 22;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 7 QRPDPQGR 15  
Db 48 QRPDPDPGK 56

Search completed: December 3, 2002, 14:26:26  
Job time: 10 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 3, 2002, 14:26:02 ; Search time 27.2 Seconds  
(without alignments)  
121.204 Million cell updates/sec

Title: US-09-924-654-4\_COPY\_550\_565

Perfect score: 92

Sequence: 1 PSHKSTQRPQPGRQ 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	92	100.0	583	4 Q92802	Q92802 homo sapien
2	58	63.0	322	10 Q9LE66	Q9LE66 oryza sativ
3	52	56.5	629	2 Q07639	Q07639 streptomyc
4	52	56.5	743	5 Q9VK47	Q9VK47 drosophila
5	51	55.4	188	11 Q62106	Q62106 mus musculu
6	51	55.4	356	10 Q9LIP1	Q9LIP1 arabidopsis
7	50	54.3	210	4 Q9UJ89	Q9UJ89 homo sapien
8	50	54.3	210	4 Q9BU80	Q9BU80 homo sapien
9	50	54.3	814	4 Q9HCW7	Q9HCW7 dictyosteli
10	50	54.3	898	5 Q8TIG7	Q8TIG7 dictyosteli
11	49.5	53.8	615	11 Q63614	Q63614 rattus norv
12	49	53.3	186	11 Q99PU6	Q99PU6 rattus norv
13	49	53.3	359	5 Q9XZT0	Q9XZT0 drosophila
14	49	53.3	379	13 Q91810	Q91810 xenopus lae
15	49	53.3	440	4 Q8TF74	Q8TF74 homo sapien
16	49	53.3	440	4 Q8TE44	Q8TE44 homo sapien

17	49	53.3	708	5 Q8SX98	Q8SX98 drosophila
18	49	53.3	708	10 Q9SX31	Q9SX31 arabidopsis
19	49	53.3	828	5 Q9W292	Q9W292 drosophila
20	49	53.3	901	4 Q60391	Q60391 homo sapien
21	49	53.3	903	10 Q9PFR5	Q9PFR5 oryza sativ
22	48.5	52.7	952	10 Q8S920	Q8S920 oryza sativ
23	48	52.2	256	10 Q8RUE4	Q8RUE4 oryza sativ
24	48	52.2	453	5 Q9U903	Q9U903 asterina pe
25	48	52.2	534	11 Q99LH7	Q99LH7 mus musculu
26	48	52.2	1353	3 Q74162	Q74162 coprinus ci
27	48	52.2	1413	5 Q9VJJ8	Q9VJJ8 drosophila
28	48	52.2	1424	5 Q9VUJ9	Q9VUJ9 drosophila
29	48	52.2	2289	3 Q9HFW4	Q9HFW4 ustiliago ma
30	47.5	51.6	370	5 Q25650	Q25650 plasmodium
31	47	51.1	259	10 Q9FQ22	Q9FQ22 oryza sativ
32	47	51.1	329	10 Q9ZTB8	Q9ZTB8 zea mays (m
33	47	51.1	345	12 Q66541	Q66541 human herpe
34	47	51.1	457	5 Q961D0	Q961D0 drosophila
35	47	51.1	472	10 Q8S5P9	Q8S5P9 oryza sativ
36	47	51.1	564	10 Q9SN05	Q9SN05 arabidopsis
37	47	51.1	575	11 Q91ZG4	Q91ZG4 mus musculu
38	47	51.1	600	5 Q8T2G2	Q8T2G2 dictyosteli
39	47	51.1	647	5 Q9VC60	Q9VC60 drosophila
40	47	51.1	661	10 Q9AWM4	Q9AWM4 oryza sativ
41	47	51.1	671	3 Q94113	Q94113 pneumocysti
42	47	51.1	749	5 Q9BIW6	Q9BIW6 drosophila
43	47	51.1	749	5 Q9W384	Q9W384 drosophila
44	47	51.1	779	5 Q01299	Q01299 caenorhabdi
45	47	51.1	859	5 Q95QH5	Q95QH5 caenorhabdi

## ALIGNMENTS

### RESULT 1

Q92802	PRELIMINARY;	PRT;	583	AA.
ID	Q92802	PRELIMINARY;	PRT;	583
AC	Q92802			
DT	01-FEB-1997 (TREMBLrel. 02, Created)			
DT	01-FEB-1997 (TREMBLrel. 02, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	Hypothetical 67.5 kDa protein.			
OS	Homo sapiens (Human)			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RP	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=96411650; PubMed=8812419;			
RA	Couch F.J., Rommens J.M., Neuhausen S.L., Belanger C., Dumont M.,			
RA	Kenneth A., Bell R., Berry S., Bogden R., Cannon-Albright L.,			
RA	Farid L., Frye C., Hattier T., Janecki T., Jiang P., Kehrer R.,			
RA	LeBlanc J.F., McArthur-Morrison J., McSweeney D., Miki Y., Peng Y.,			
RA	Samson C., Schroeder M., Snyder S.C., Stringfellow M., Stroup C.,			
RA	Swedlund B., Swensen J., Teng D., Thakur S., Tran T., Tranchant M.,			
RA	Welver-Feldhaus J., Wong A.K.C., Shizuya H., Labrie F., Skolnick M.H.,			
RA	Goldgar D.E., Kamb A., Weber B.L., Tavcigian S.V., Simard J.;			
RT	"Generation of an integrated transcription map of the BRCA2 region on			
RL	chromosome 13q12-q13."			
RL	Genomics 36:86-99(1996).			
RP	[2]			
RP	SEQUENCE FROM N.A.			
RA	Rhodes S.;			
RL	Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.			
RP	[3]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=COLON;			
RA	Strausberg R.;			
RL	Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; U50532; AAC50875.1; -.			
DR	EMBL; AL049783; CAB42441.1; -.			
DR	EMBL; BC010643; AAH10643.1; -.			
KW	Hypothetical protein.			

SQ SEQUENCE 583 AA; 67459 MW; 318DC4D81CD0FF2A CRC64;  
 Query Match 100.0%; Score 92; DB 4; Length 583;  
 Best Local Similarity 100.0%; Pred. No. 4.9e-06;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PSHKSTORPPPPQGRQ 16  
 |||||  
 DB 550 PSHKSTORPPPPQGRQ 565  
 |||||

RESULT 2  
 ID Q9LE66 PRELIMINARY; PRT; 322 AA.  
 AC Q9LE66;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)  
 DE Hypothetical protein.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Eriactoidae; Oryzeae; Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. NIPPONBARE;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nippobare (GA3) genomic DNA, chromosome 1, PAC  
 clone:PO466B10.";  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. NIPPONBARE;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nippobare (GA3) genomic DNA, chromosome 1, PAC  
 clone:PO031E09.";  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AP002093; BAA96179.1;  
 DR EMBL; AP002092; BAA96137.1;  
 SQ SEQUENCE 322 AA; 36865 MW; 1134E31A1056B9F3 CRC64;

Query Match 63.0%; Score 58; DB 10; Length 322;  
 Best Local Similarity 64.3%; Pred. No. 0.38;  
 Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 PSHKSTORPPPPQ 14  
 |||||  
 DB 250 PAHGERQRPPEPRG 263  
 |||||

RESULT 3  
 ID Q07639 PRELIMINARY; PRT; 629 AA.  
 AC Q07639;  
 DT 01-JUN-1998 (TrEMBLrel. 06, Created)  
 DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Membrane translocator.  
 GN AMFB.  
 OS Streptomyces griseus.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=1911;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=93209944; PubMed=8458843;  
 RA Ueda K., Miyake K., Horinouchi S., Beppu T.;  
 RT "A gene cluster involved in aerial mycelium formation in Streptomyces  
 griseus encodes proteins similar to the response regulators of two-  
 component regulatory systems and membrane translocators.";  
 RL J. Bacteriol. 175:2006-2016(1993).  
 RN [2]

RP SEQUENCE FROM N.A.  
 RC STRAIN=HH1;  
 RX MEDLINE=98422460; PubMed=9748440;  
 RA Ueda K., Hsieh C.-W., Tosaki T., Shinkawa H., Beppu T., Horinouchi S.;  
 RT "Characterization of an A-factor-responsive repressor for amfR  
 essential for onset of aerial mycelium formation in Streptomyces  
 griseus.";  
 RL J. Bacteriol. 180:5085-5093(1998).  
 CC -I- FUNCTION: INVOLVED IN AERIAL MYCELIUM FORMATION.  
 DR EMBL; AB06206; BAA33538.1;  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR001140; ABC\_tranprtm.  
 DR InterPro; IPR003439; ABC\_transportr.  
 DR Pfam; PF00664; ABC\_membrane; 1.  
 DR Pfam; PF00005; ABC\_tran; 1.  
 DR ProDom; PD000006; ABC\_transportr; 1.  
 DR SMART; SM00382; AAA\_1; ABC\_TRANSPORTER; 1.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER; 1.  
 KW Membrane.  
 SQ SEQUENCE 629 AA; 65170 MW; 878450C46AF9744B CRC64;

Query Match 56.5%; Score 52; DB 2; Length 629;  
 Best Local Similarity 50.0%; Pred. No. 5.8;  
 Matches 10; Conservative 3; Mismatches 3; Indels 4; Gaps 1;

QY 1 PSHKSTORPP-PPQGRQ 16  
 |||||  
 DB 604 PRHPTTORPPQAPPPKSR 623  
 |||||

RESULT 4  
 ID Q9VK47 PRELIMINARY; PRT; 743 AA.  
 AC Q9VK47;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE CG5075 protein.  
 GN CG5075.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Flesler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kianos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svitek R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Massaman D.A., Weinstock G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhang G., Zhao Q., Zheng L.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
RT "The genome sequence of Drosophila melanogaster";  
RL Science 287:2185-2195(2000).  
CC -!- SIMILARITY: BELONGS TO THE ATPASE ALPHA/BETA CHAINS FAMILY.  
DR EMBL; AF003638; AAF53234.1; -;  
DR FlyBase; FBgn0032464; CG5075.  
DR InterPro; IPR000793; ATPase\_a/bC.  
DR InterPro; IPR000194; ATPase\_a/bCentre.  
DR InterPro; IPR004100; ATPase\_a/bN.  
DR Pfam; PF00006; ATP-synt\_ab; 1.  
DR Pfam; PF00306; ATP-synt\_ab\_C; 1.  
DR Pfam; PF02874; ATP-synt\_ab\_N; 1.  
DR TIGRFAM; TIGR01042; V-ATPase\_V1\_A; 1.  
DR PROSITE; PS00152; ATPASE ALPHA/BETA; 1.  
KW Hydrogen ion transport; Hydrolase  
SQ SEQUENCE 743 AA; 82307 MW; 357E9504876BCFD3 CRC64;  
Query Match 56.5%; Score 52; DB 5; Length 743;  
Best Local Similarity 64.3%; Pred. No. 6.8;  
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
QY 3 HKSTQRPDPGGRQ 16  
DB 61 HKSNVPVPPQSRK 74  
RESULT 5  
Q62106 PRELIMINARY; PRT; 188 AA.  
AC Q62106;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE Proline-rich salivary protein (fragment).  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=8603799; PubMed=3840480;  
RA Clements S.; Mehansho H.; Carlson D.M.;  
RT "Novel multigene families encoding highly repetitive peptide  
RT sequences: Sequence analyses of rat and mouse proline-rich protein  
RT cDNAs";  
RL J. Biol. Chem. 260:13471-13477(1985).  
DR EMBL; M19419; AAA40002.1; -;  
DR InterPro; IPR002965; P rich extensn.  
DR PRINTS; PR01217; PRICHEXTENS.  
FT NON TER 1  
SQ SEQUENCE 188 AA; 18657 MW; A6598BC5560473FE CRC64;  
Query Match 55.4%; Score 51; DB 11; Length 188;  
Best Local Similarity 56.2%; Pred. No. 2.6;  
Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;  
QY 1 PSKSTQRPDPGGRQ 16  
DB 69 PQORSPQSPPPGPGQ 84  
RESULT 6  
Q9LIP1 PRELIMINARY; PRT; 356 AA.  
AC Q9LIP1;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
GB|AAD31058.1.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=COLUMBIA;  
RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=COLUMBIA;  
RX MEDLINE=20363099; PubMed=10907853;  
RA Nakamura Y.;  
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.  
RT Sequence features of the regions of 4,251,695 bp covered by ninety Pl,  
RT TAC and BAC clones";  
RL DNA Res. 7:217-221(2000).  
DR EMBL; AP001298; BAB02195.1; -;  
DR InterPro; IPR002965; P-rich\_extensn.  
DR InterPro; IPR001865; Ribosomal\_S2.  
DR PRINTS; PR01217; PRICHEXTENS.  
DR PROSITE; PS00962; RIBOSOMAL S2 1; UNKNOWN 1.  
SQ SEQUENCE. 356 AA; 39906 MW; 124AB52ADE8C08EF CRC64;  
Query Match 55.4%; Score 51; DB 10; Length 356;  
Best Local Similarity 56.2%; Pred. No. 4.7;  
Matches 9; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
QY 1 PSKSTQRPDPGGRQ 16  
DB 112 PEDNQQRPPQAPQ 127  
RESULT 7  
Q9UJ89 PRELIMINARY; PRT; 210 AA.  
AC Q9UJ89;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)  
DE LAGE-1L protein.  
GN LAGE-1L.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX TISSUE=MELANOMA;  
RC MEDLINE=99325550; PubMed=10399963;  
RA Aarnoudse C.A.; Van den Doel P.B.; Heemskerk B.; Schrier P.I.;  
RT "Interleukin-2-induced, melanoma-specific T cells recognize CAMEL, an  
RT unexpected translation product of LAGE-1";  
RL Int. J. Cancer 82:442-448(1999).  
DR EMBL; AJ012835; CAA10196.1; -;  
SQ SEQUENCE 210 AA; 21060 MW; 1DD0B1829735B60A CRC64;  
Query Match 54.3%; Score 50; DB 4; Length 210;  
Best Local Similarity 54.5%; Pred. No. 4;  
Matches 12; Conservative 1; Mismatches 3; Indels 6; Gaps 2;  
QY 1 PSKSTQRPDPGGRQ 16  
DB 171 PSKSTQRPDPGGRQ 192

```

DE Hypothetical 101.8 kDa protein.
OC Dictyostelium discoideum (Slime mold).
OS Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.,
RT "Sequence and Analysis of Chromosome 2 of Dictyostelium."
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC16548; AAL99323.1; -.
KW Hypothetical protein.
SQ SEQUENCE 898 AA; 101823 MW; AE8B1512BB64A148 CRC64;

Query Match 54.3%; Score 50; DB 5; Length 898;
Best Local Similarity 56.2%; Pred. No. 16;
Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 PSHKSTQRP PPPQGRQ 16
   |||:|||||:|
DB 149 PPSQSSQPPPPQQQQ 164

RESULT 11
Q63614 PRELIMINARY; PRT; 615 AA.
ID O63614 AC
Q63614,
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TEMBLrel. 20, Last annotation update)
DE Tyrosine kinase receptor (Fragment).
DE Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OC NCBI_TaxID=101116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94043123; PubMed=7693687;
RA Benhamou M., Ryba N.J.P., Nishikata H., Kihara H., Siraganian R.P.;
RT "protein tyrosine kinase p72yck in high affinity IGE receptor
RT signaling: identification as a component of pp72 and association with
RT the gamma chain of the receptor after receptor aggregation.";
RL J. Biol. Chem. 268;23318-23324(1993).
DR EMBL; L20838; AAA42308.1; -.
DR HSSP; P43405; 1A81.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR000980; SH2.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00017; SH2; 2.
DR PRINTS; PR00401; SH2DOMAIN.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR ProDom; PD000093; SH2; 2.
DR SMART; SM00252; SH2; 2.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1..
DR PROSITE; PS50011; PROTEIN KINASE DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TVR; 1.
DR PROSITE; PS50001; SH2; 2.
KW Kinase; Receptor; Tyrosine-protein kinase.
FT NON_TER 1
FT NON_TER 615
SQ SEQUENCE 615 AA; 70008 MW; B41D4CEE69A567A2 CRC64;

Query Match 53.8%; Score 49.5; DB 11; Length 615;
Best Local Similarity 64.3%; Pred. No. 14;
Matches 9; Conservative 1; Mismatches 1; Indels 3; Gaps 1;

QY 1 PSHKSTQRP PPPPG 14

```

```

Db 281 PGHK---KPPPOG 291
||| :|||||
1 1 PSKSTQRPPOGR 15
||| :|||||
144 PSPKSEITPPPOPR 158

RESULT 13
Q9XZT0 PRELIMINARY; PRT; 359 AA.
AC Q9XZT0; Q9W5D1;
DT 01-NOV-1999 (T-EMBLrel. 12, Created)
DT 01-NOV-1999 (T-EMBLrel. 12, Last sequence update)
DT 01-JUN-2000 (T-EMBLrel. 14, Last annotation update)
DE EG:34F3.10 protein (CG13358 protein)
DN EG:34F3.10 OR CG13358.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY.
RX MEDLINE=20196006; PubMed=107311132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers J.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

Db 281 PGHK---KPPPOG 291
||| :|||||
1 1 PSKSTQRPPOGR 15
||| :|||||
144 PSPKSEITPPPOPR 158

RESULT 12
Q99PU6 PRELIMINARY; PRT; 186 AA.
AC Q99PU6;
DT 01-JUN-2001 (T-EMBLrel. 17, Created)
DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)
DT 01-WAR-2002 (T-EMBLrel. 20, Last annotation update)
DE Branched-chain alpha-keto acid dihydrolipoyl acyltransferase
DE (Fragment)
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21099342; PubMed=11168412;
RA Ono K., Hatozaki M., Suzuki T., Mori T., Hata H., Kochi H.;
RT "cDNA cloning of the chicken branched-chain a-keto acid dehydrogenase
RT complex: Chicken-specific residues of the acyltransferase affect the
RT overall activity and the interaction with the dehydrogenase.";
RL Eur. J. Biochem. 268:727-736(2001).
CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL (BY SIMILARITY).
CC -!- SIMILARITY: CONTAINS 1 LIPOYL-BINDING DOMAIN.
DR EMBL; AB047915; BAB32668.1; -.
DR HSSP; P11961; 2PDD.
DR InterPro; IPR004167; E3 binding lipoyl.
DR InterPro; IPR004167; E3 binding lipoyl.
DR InterPro; IPR003016; Lipoyl.
DR Pfam; PF00364; biotin lipoyl; 1.
DR Pfam; PF02817; e3 binding; 1.
DR PROSITE; PS00189; LIPOYL; 1.
KW Acyltransferase; Lipoyl; Mitochondrion; Transferase.
FT NON_TER 1
FT NON_TER 186
SQ SEQUENCE 186 AA; 20780 MW; 39214AF36DC45B54 CRC64;

Query Match 53.3%; Score 49; DB 11; Length 186;
Best Local Similarity 66.7%; Pred. No. 5.1;
Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 PSKSTQRPPOGR 15
||| :|||||
144 PSPKSEITPPPOPR 158

RESULT 13
Q9XZT0 PRELIMINARY; PRT; 359 AA.
AC Q9XZT0; Q9W5D1;
DT 01-NOV-1999 (T-EMBLrel. 12, Created)
DT 01-NOV-1999 (T-EMBLrel. 12, Last sequence update)
DT 01-JUN-2000 (T-EMBLrel. 14, Last annotation update)
DE EG:34F3.10 protein (CG13358 protein)
DN EG:34F3.10 OR CG13358.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY.
RX MEDLINE=20196006; PubMed=107311132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers J.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

```

```

RA Abril J.F., Agbayani A., An H.-J., Andrews-Efannkoch C., Baldwin D.,
RA Balwale R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.W., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=OREGON-R;
RA Valenti P., Salles C., Campbell L., Glover D.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003418; AAF45553.1; ALT_INIT.
DR EMBL; AL031583; CAB41346.1; -.
DR Flybase; FBgn0026874; EG:34F3.10.
SQ SEQUENCE 359 AA; 38488 MW; 083A6CF2C6F1D74E CRC64;

Query Match 53.3%; Score 49; DB 5; Length 359;
Best Local Similarity 66.7%; Pred. No. 9.6;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 PSKSTQRPPOGR 12
||| :|||||
161 PADASTRRPPPP 172

Db 161 PADASTRRPPPP 172

RESULT 14
Q91810 PRELIMINARY; PRT; 379 AA.
AC Q91810;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE Proline rich protein.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=OOCYTE;
RA Nishimatsu S.I., Satoshi, Oda, Naoto, Ueno;
RL Submitted (AUG-1992) to the EMBL/GenBank/DBJ databases.
DR EMBL; X68249; CAA48321.1; -.
DR InterPro; IPR003124; WH2.
DR Pfam; PF02205; WH2; 1.
DR PRINTS; PR01574; TUBBYPROTEIN.

```

DR SMART; SM00246; WH2; 1.  
SQ SEQUENCE 379 AA; 38859 MW; 84DC0FB24F971AAD CRC64;

Query Match 53.3%; Score 49; DB 13; Length 379;  
Best Local Similarity 57.1%; Pred. No. 10;  
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 HKSTQRPDPGGRQ 16  
| : |||||  
Db 178 HSSAPPPPPGRR 191

## RESULT 15

Q8TF74

ID Q8TF74 PRELIMINARY; PRT; 440 AA.

AC Q8TF74;

DT 01-JUN-2002 (TrEMBLrel. 21, Created)

DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)

DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

DR Putative crl6 and wip like protein.

OS Homo sapiens (Human)

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA Kato M., Miki H., Takenawa T.;

RT "Homo sapiens mRNA for putative crl6 and wip like protein.";

RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AB043786; BAB85113.1; -

SQ SEQUENCE 440 AA; 46288 MW; 7386B44F995AD798 CRC64;

Query Match

53.3%; Score 49; DB 4; Length 440;

Best Local Similarity 57.1%; Pred. No. 12;

Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 HKSTQRPDPGGRQ 16

| : |||||

Db 172 HSSAPPPPPGRR 185

Search completed: December 3, 2002, 14:27:23

Job time : 29.2 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 3, 2002, 14:26:06 ; Search time 12.2667 Seconds  
(without alignments)  
38.378 Million cell updates/sec

Title: US-09-924-654-4\_COPY\_550\_565

Perfect score: 92

Sequence: 1 PSHKSTQRPPPPQGRQ 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	92	100.0	583	2	US-08-616-392C-4
2	50	54.3	210	2	US-08-791-495-5
3	48	52.2	2289	3	US-09-051-019-2
4	47	51.1	17	4	US-08-602-999A-398
5	47	51.1	17	4	US-09-500-124-398
6	47	51.1	345	1	US-08-031-148-4
7	47	51.1	345	3	US-08-415-838-4
8	47	51.1	345	4	US-09-205-169-4
9	45	48.9	322	4	US-09-383-586-33
10	44.5	48.4	70	4	US-09-006-428A-15
11	44.5	48.4	351	4	US-08-466-465-6
12	44.5	48.4	472	2	US-08-742-621-5
13	44.5	48.4	472	4	US-08-842-079-15
14	44.5	48.4	986	4	US-09-403-618A-5
15	44	47.8	17	4	US-08-602-999A-399
16	44	47.8	17	4	US-09-500-124-399
17	44	47.8	61	4	US-09-314-268-142
18	44	47.8	466	4	US-08-580-031A-16
19	43.5	47.3	447	4	US-09-191-608-19
20	43.5	47.3	471	4	US-09-191-608-17
21	43.5	47.3	497	4	US-09-191-608-20
22	43	46.7	491	4	US-09-930-872-2
23	43	46.7	962	4	US-09-442-100-6
24	43	46.7	1130	4	US-09-442-100-4
25	43	46.7	1224	4	US-09-930-872-4
26	42	45.7	23	1	US-08-268-251-51
27	42	45.7	23	5	PCT-US93-01112-51

28	42	45.7	34	4	US-09-596-938-1	Sequence 1, Appli
29	42	45.7	112	1	US-08-446-919A-5	Sequence 5, Appli
30	42	45.7	328	1	US-08-414-926A-9	Sequence 9, Appli
31	42	45.7	328	2	US-08-926-922-9	Sequence 9, Appli
32	42	45.7	328	3	US-09-253-682-9	Sequence 9, Appli
33	42	45.7	328	4	US-09-527-657-9	Sequence 2, Appli
34	42	45.7	346	4	US-08-894-139-2	Sequence 9, Appli
35	42	45.7	437	2	US-08-935-450-8	Sequence 8, Appli
36	42	45.7	696	3	US-08-906-865-4	Sequence 4, Appli
37	42	45.7	696	4	US-09-129-668-4	Sequence 2, Appli
38	42	45.7	1487	2	US-08-760-489-2	Sequence 4, Appli
39	42	45.7	1487	2	US-08-760-489-4	Sequence 2, Appli
40	42	45.7	1487	4	US-09-185-373-2	Sequence 4, Appli
41	42	45.7	1487	4	US-09-185-373-4	Sequence 7, Appli
42	42	45.7	2842	1	US-07-741-940-7	Sequence 7, Appli
43	42	45.7	2842	1	US-08-289-548A-7	Sequence 7, Appli
44	42	45.7	2842	1	US-08-452-654-7	Sequence 7, Appli
45	42	45.7	2842	4	US-08-449-731-7	Sequence 7, Appli

#### ALIGNMENTS

#### RESULT 1

US-08-616-392C-4

; Sequence 4, Application US/08616392C

; Patent No. 5998165

; GENERAL INFORMATION:

; APPLICANT: Goold, Richard D.

; APPLICANT: Akerblom, Ingrid E.

; APPLICANT: Seilhamer, Jeffrey

; APPLICANT: Coleman, Roger

; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES PANCI

; TITLE OF INVENTION: AND PANCI ASSOCIATED WITH PANCREATIC CANCER

; NUMBER OF SEQUENCES: 12

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Drive

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/616,392C

; FILING DATE: 15-MAR-1996

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/581,240

; FILING DATE: 29-DEC-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Billings, Lucy J.

; REGISTRATION NUMBER: 36,749

; REFERENCE/DOCKET NUMBER: PP-0052-1US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 650-855-0555

; TELEFAX: 650-845-4166

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 583 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; IMMEDIATE SOURCE:

; LIBRARY: hnt

; CLONE: 496071

US-08-616-392C-4

Query Match 100.0%; Score 92; DB 2; Length 583;

Best Local Similarity 100.0%; Pred. No. 1.1e-05;



Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PSHKSTORPPPPQGRQ 16

Db 550 PSHKSTORPPPPQGRQ 565

## RESULT 2

US-08-791-495-5

Sequence 5, Application US/08791495

Patent No. 581519

GENERAL INFORMATION:

APPLICANT: Leth, Bernard

APPLICANT: Lucas, Sophie

APPLICANT: De Smet, Charles

APPLICANT: Godelaine, Daniele

APPLICANT: Boon-Falleur, Thierry

TITLE OF INVENTION: LL-1 TUMOR SPECIFIC GENES

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:

ADDRESSEE: Wolf, Greenfield & Sacks, P.C.

STREET: 600 Atlantic Avenue

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02210

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/791.495

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Van Amsterdam, John R.

REGISTRATION NUMBER: 40,212

REFERENCE/DOCKET NUMBER: L0461/7005

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-720-3500

TELEFAX: 617-720-2441

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 210 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-791-495-5

Query Match 54.3%; Score 50; DB 2; Length 210;

Best Local Similarity 54.5%; Pred. No. 3.5;

Matches 12; Conservative 1; Mismatches 3; Indels 6; Gaps 2;

Qy 1 PSHK-STOR-----PPPPQGRQ 16

Db 171 PKHKVSEORPGTGPPEGAQ 192

## RESULT 3

US-09-051-019-2

Sequence 2, Application US/09051019

Patent No. 6103229

GENERAL INFORMATION:

APPLICANT: KAHMANN, Regine and QUABECK-SIEGER, Claudia

TITLE OF INVENTION: Regulatory gene from Ustilago maydis

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: Keil & Weinkauff

STREET: 1101 Connecticut Avenue

CITY: Washington

STATE: D.C.

COUNTRY: USA

US-08-602-999A-398

ZIP: 20036

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.5 inch, 1.2 Mb storage

COMPUTER: IBM AT-compatible, Pentium processor

OPERATING SYSTEM: Windows 98

SOFTWARE: WordPerfect version 6.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/051.019

FILING DATE: 31-MAR-1998

CLASSIFICATION:

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 2289 amino acids

TYPE: amino acid

TOPOLOGY: linear

US-09-051-019-2

Query Match 52.2%; Score 48; DB 3; Length 2289;

Best Local Similarity 60.0%; Pred. No. 78;

Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 1 PSHKSTORPPPPQGR 15

Db 201 PTHASTPRCPPLRLR 215

## RESULT 4

US-08-602-999A-398

Sequence 398, Application US/08602999A

Patent No. 6184205

GENERAL INFORMATION:

APPLICANT: SPARKS, Andrew B.

APPLICANT: KAY, Brian K.

APPLICANT: THORN, Judith M.

APPLICANT: QUILLIAM, Lawrence A.

APPLICANT: DER, Channing J.

APPLICANT: FOWLKES, Dana M.

APPLICANT: RIDER, James E.

TITLE OF INVENTION: SH3 BINDING PEPTIDES AND METHODS OF

ISOLATING AND USING SAME

NUMBER OF SEQUENCES: 467

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/602.999A

FILING DATE: 16-FEB-1996

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Misrock, S. Leslie

REGISTRATION NUMBER: 18,872

REFERENCE/DOCKET NUMBER: 1101-202

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 790-9090

TELEFAX: (212) 869-9741/8864

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 398:

SEQUENCE CHARACTERISTICS:

LENGTH: 17 amino acids

TYPE: amino acid

TOPOLOGY: unknown

MOLECULE TYPE: peptide

US-08-602-999A-398

APPLICANT: Middelcorp, Jaap Michiel.  
TITLE OF INVENTION: Peptides and nucleic acid sequences  
TITLE OF INVENTION: related to the Epstein-Bar virus.  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Akzo Pharma  
STREET: 1330-A Piccard Drive  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850-4377

COUNTRY: USA  
ZIP: 20850-4377  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/031.148  
FILING DATE: 19930312  
CLASSIFICATION: 530

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/031.148  
FILING DATE: 19930312  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: EP 92200731.6  
FILING DATE: 11-MAR-1992

CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: EP 9200721.6  
FILING DATE: 13-MAR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Bobrowicz, Donna  
REGISTRATION NUMBER: 32 196

FILING DATE: 13-MAR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Bobrowicz, Donna  
REGISTRATION NUMBER: 32,196  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 345 amino acids

REGISTRATION NUMBER: 32,196  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
   LENGTH: 345 amino acids  
   TYPE: AMINO ACID  
   STRANDEDNESS: single  
   TOPOLOGY: linear

LENGTH: 345 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
ORIGINAL SOURCE:  
ORGANISM: Epstein-Barr virus

TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 ORIGINAL SOURCE: Epstein-Barr virus  
 ORGANISM: Epstein-Barr virus  
 8-031-148-4  
 serv Match 51.1% Score 47 DB 1: Le

ORGANISM: Epstein-Barr virus  
8-031-148-4

Very Match	51.1%;	Score 47;	DB 1;	Le
1st Local Similarity	53.3%;	Pred. NO. 16;		
atches	8;	Conservative	2;	Mismatches 5;

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	
1	2	3	4	5	6	7	8	9	10	11																																																																																										

2 SHKSTQRPppPQGRQ.16  
|||:|||||  
248 SHQAAQPPPPGTQ 262

248 SHQAAQAPPPPGTQ 262

LT 7  
8-415-838-4  
quence 4, Application US/08415838  
tent No. 608327  
GENERAL INFORMATION:  
APPLICANT: Middelorp, Jaap Michiel.  
TITLE OF INVENTION: Peptides and nucleic acid s

tent No. 6008327  
GENERAL INFORMATION:  
APPLICANT: Middelorp, Jaap Michiel.  
TITLE OF INVENTION: Peptides and nucleic acid sequences  
TITLE OF INVENTION: related to the Epstein-Barr  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:

ADDRESS: Akzo-No. 6008327el Patent Department  
 STREET: 1300 Piccard Drive, Suite 206  
 CITY: Rockville

CORRESPONDENCE ADDRESS:  
ADDRESSEE: AKZO-No. 6008327el Patent Department  
STREET: 1300 Piccard Drive, Suite 206  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850

CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.2  
CURRENT APPLICATION DATA:

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.22  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/415,838  
FILING DATE: 03-APR-1995  
PRIOR APPLICATION DATA:

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/415,838  
FILING DATE: 03-APR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: EP 92200721.6  
FILING DATE: 13-MAR-1992

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: EP 92200721.6  
FILING DATE: 13-MAR-1992

ATTORNEY/AGENT INFORMATION:  
NAME: Gormley, Mary E.  
REGISTRATION NUMBER: 34,409  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 345 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
ORIGINAL SOURCE:  
ORGANISM: Epstein-Barr virus  
US-08-415-838-4

Query Match 51.1%; Score 47; DB 3; Length 345;  
Best Local Similarity 53.3%; Pred. No. 16;  
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 SHKSTORPPPPQGRQ 16  
||:| |||||  
DB 248 SHOAAQAQPPPPGTQ 262

## RESULT 8

US-09-205-169-4

; Sequence 4, Application US/09205169

; Patent No. 6365717

; GENERAL INFORMATION:

; APPLICANT: Middeldorp, Jaap Michiel.

; TITLE OF INVENTION: Peptides and nucleic acid sequences  
related to the Epstein-Barr virus.

; NUMBER OF SEQUENCES: 22

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Akzo-No. 6365717el Patent Department

; STREET: 1300 Piccard Drive, Suite 206

; CITY: Rockville

; STATE: Maryland

; COUNTRY: USA

; ZIP: 20850

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/205,169

; FILING DATE: 04-Dec-1998

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/415,838

; FILING DATE: <Unknown>

; ATTORNEY/AGENT INFORMATION:

; NAME: Gormley, Mary E.

; REGISTRATION NUMBER: 34,409

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 345 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

; ORIGINAL SOURCE:

; ORGANISM: Epstein-Barr virus

; SEQUENCE DESCRIPTION: SEQ ID NO: 4:

US-09-205-169-4

Query Match 51.1%; Score 47; DB 4; Length 345;

Best Local Similarity 53.3%; Pred. No. 16;

Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 SHKSTORPPPPQGRQ 16  
||:| |||||  
DB 248 SHOAAQAQPPPPGTQ 262

## RESULT 9

US-09-383-586-33

; Sequence 33, Application US/09383586

; Patent No. 6242419

; GENERAL INFORMATION:

; APPLICANT: Strachan, Lorna

; APPLICANT: Sleeman, Matthew

; APPLICANT: Abernethy, Nevin

; APPLICANT: Onrust, Rene

; APPLICANT: Kumbie, Anand

; APPLICANT: Murison, Greg

; TITLE OF INVENTION: Compounds isolated from stromal cells

; FILE REFERENCE: 11000.1037c1

; CURRENT APPLICATION NUMBER: US/09/383,586

; CURRENT FILING DATE: 1999-08-26

; NUMBER OF SEQ ID NOS: 38

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 33

; LENGTH: 322

; TYPE: PRT

; ORGANISM: Human

US-09-383-586-33

Query Match 48.9%; Score 45; DB 4; Length 322;

Best Local Similarity 56.2%; Pred. No. 28;

Matches 9; Conservative 1; Mismatches 2; Indels 4; Gaps 1;

QY 1 PSHKSTORPPPPQGRQ 16  
||:| |||||

DB 17 PSH----RPPPEAPQ 28

## RESULT 10

US-09-006-428A-15

; Sequence 15, Application US/09006428A

; Patent No. 6444439

; GENERAL INFORMATION:

; APPLICANT: Jing Li

; APPLICANT: Kazuhisa Nishizawa

; APPLICANT: Wenglian An

; APPLICANT: Ellis L. Reinherz

; TITLE OF INVENTION: CLONING AND CHARACTERIZATION OF A

FILE REFERENCE: 1062.1020-000

; CURRENT APPLICATION NUMBER: US/09/006,428A

; CURRENT FILING DATE: 1998-01-13

; NUMBER OF SEQ ID NOS: 28

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 15

; LENGTH: 70

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-006-428A-15

Query Match 48.4%; Score 44.5; DB 4; Length 70;

Best Local Similarity 50.0%; Pred. No. 7;

Matches 9; Conservative 2; Mismatches 4; Indels 3; Gaps 1;

QY 1 PSHKS--TORPPPPQGR 15  
||:| |||||

DB 5 PGRSQAPSHRPPPPGHR 22

## RESULT 11

US-08-466-465-6

; Sequence 6, Application US/08466465

; Patent No. 6162432

; GENERAL INFORMATION:

; APPLICANT: Wallner, Barbara P.

; APPLICANT: Cooper, Kevin D.

; TITLE OF INVENTION: Method of Prophylaxis or Treatment of Antigen

```

; TITLE OF INVENTION: Presenting Cell Driven Skin Conditions Using
; TITLE OF INVENTION: Inhibitors of the CD2/LFA-3 Interaction
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, Suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/466,465
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/08755
; FILING DATE: 06-OCT-1992
; APPLICATION DATA:
; APPLICATION NUMBER: US 07/862,022
; FILING DATE: 12-APR-1992
; APPLICATION DATA:
; APPLICATION NUMBER: US 07/770,969
; FILING DATE: 07-OCT-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis (PLM)
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: BGP-111CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 351 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-466-465-6

Query Match 48.4%; Score 44.5; DB 4; Length 351;
Best Local Similarity 50.0%; Pred. No. 36;
Matches 9; Conservative 2; Mismatches 4; Indels 3; Gaps 1;

QY 1 PSHKS---TORPPPPQGR 15
| | | | | | | | | |
Db 286 PGRSQAPSHRPPPPGHR 303

RESULT 12
US-08-742-621-5
; Sequence 5, Application US/08742621
; Patent No. 5856129
; GENERAL INFORMATION:
; APPLICANT: HILLMAN, JENNIFER L.
; APPLICANT: COLEMAN, ROGER
; TITLE OF INVENTION: NOVEL HUMAN PURINOCEPTOR
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: IBM Compatible
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/742,621
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0147 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 472 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 558831
; US-08-742-621-5

Query Match 48.4%; Score 44.5; DB 2; Length 472;
Best Local Similarity 60.0%; Pred. No. 49;
Matches 9; Conservative 1; Mismatches 4; Indels 1; Gaps 1;

QY 1 PSHKSTORPP-PPQG 14
| | | | | | | | | |
Db 395 PSHYSQDQPPSPSG 409

RESULT 13
US-08-842-079-15
; Sequence 15, Application US/08842079
; Patent No. 6133434
; GENERAL INFORMATION:
; APPLICANT: BUELL, GARY N.
; APPLICANT: SUPRENTANT, ANNMARIE
; APPLICANT: KAWASHIMA, ERIC
; TITLE OF INVENTION: A PURINERGIC RECEPTOR
; FILE REFERENCE: 1430-160
; CURRENT APPLICATION NUMBER: US/08/842,079
; CURRENT FILING DATE: 1997-04-28
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 15
; LENGTH: 472
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; US-08-842-079-15

Query Match 48.4%; Score 44.5; DB 4; Length 472;
Best Local Similarity 60.0%; Pred. No. 49;
Matches 9; Conservative 1; Mismatches 4; Indels 1; Gaps 1;

QY 1 PSHKSTORPP-PPQG 14
| | | | | | | | | |
Db 395 PSHYSQDQPPSPSG 409

RESULT 14
US-09-403-618A-5
; Sequence 5, Application US/09403618A
; Patent No. 6429358
; GENERAL INFORMATION:
; APPLICANT: BROGLIE, KAREN E.
; TITLE OF INVENTION: CORN FULLULANASE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
```

ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY  
STREET: 1007 MARKET STREET  
CITY: WILMINGTON  
STATE: DELAWARE  
COUNTRY: U.S.A.  
ZIP: 19898  
COMPUTER READABLE FORM:  
MEDIUM TYPE: DISKETTE, 3.50 INCH  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: MICROSOFT WINDOWS 95  
SOFTWARE: MICROSOFT WORD FOR WINDOWS 95 (7.0)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/403,618A  
FILING DATE: 05-No. 6429358-1999  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/045,723  
FILING DATE: MAY 6, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: MAJARIAN, WILLIAM R.  
REGISTRATION NUMBER: 41,173  
REFERENCE/DOCKET NUMBER: BB-1108  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 986 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 5:  
US-09-403-618A-5

Query Match 48.4%; Score 44.5; DB 4; Length 986;  
Best Local Similarity 60.0%; Pred. No. 1e+02; Mismatches 5; Indels 1; Gaps 1;  
Matches 9; Conservative 0;

QY 1 PSHKSTORP-PPFQG 14  
| | | | |  
Db 26 PGRSGTARPLPPQG 40

RESULT 15  
US-08-602-999A-399  
; Sequence 399, Application US/08602999A  
; Patent No. 6184205  
; GENERAL INFORMATION:  
; APPLICANT: SPARKS, Andrew B.  
; APPLICANT: KAY, Brian K.  
; APPLICANT: THORN, Judith M.  
; APPLICANT: QUILLIAM, Lawrence A.  
; APPLICANT: DER, Channing J.  
; APPLICANT: FOWLKES, Dana M.  
; APPLICANT: RIDER, James E.  
; TITLE OF INVENTION: SH3 BINDING PEPTIDES AND METHODS OF  
; TITLE OF INVENTION: ISOLATING AND USING SAME  
; NUMBER OF SEQUENCES: 467  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/602,999A  
; FILING DATE: 16-FEB-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:

NAME: Misrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 1101-202  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 399:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 17 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-602-999A-399

Query Match 47.8%; Score 44; DB 4; Length 17;  
Best Local Similarity 42.9%; Pred. No. 1.9;  
Matches 6; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 3 HKSTQRP PPPQGRQ 16  
| : : : : :  
Db 4 HPNFERKPKPKGKQ 17

Search completed: December 3, 2002, 14:29:36  
Job time : 14.2667 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 3, 2002, 14:26:31 ; Search time 7.46667 Seconds  
(without alignments)  
34.123 Million cell updates/sec

Title: US-09-924-654-4\_COPY\_550\_565

Perfect score: 92

Sequence: 1 PSHKSTQRPQPGRQ 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 102317 seqs, 15924203 residues

Total number of hits satisfying chosen parameters: 102317

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

- Database : Published Applications AA:\*
- 1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*
  - 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep:\*
  - 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*
  - 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*
  - 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*
  - 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*
  - 7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep:\*
  - 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*
  - 9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*
  - 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep:\*
  - 11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*
  - 12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep:\*
  - 13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*
  - 14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	92	100.0	583	10	US-09-924-654-4
2	92	100.0	594	10	US-09-925-300-1079
3	49	53.3	901	10	US-09-737-149-33
4	49	53.3	998	9	US-10-106-534-2
5	49	53.3	1043	10	US-09-737-149-8
6	47	51.1	196	9	US-09-989-920-224
7	47	51.1	345	9	US-10-036-729-4
8	46	50.0	80	10	US-09-864-761-41757
9	46	50.0	468	10	US-09-925-297-736
10	46	50.0	506	10	US-09-801-368-90
11	45	48.9	274	10	US-09-850-887-4
12	45	48.9	322	10	US-09-823-038A-33
13	44.5	48.4	21	10	US-09-873-106B-22
14	44.5	48.4	70	10	US-09-873-106B-11
15	44.5	48.4	351	10	US-09-796-033-6
16	44.5	48.4	351	10	US-09-730-465-6
17	44	47.8	72	10	US-09-864-761-40031
18	44	47.8	132	10	US-09-864-761-43644
19	44	47.8	466	12	US-10-095-492-16

20	44	47.8	836	10	US-09-934-323-5	Sequence 5, Appli
21	44	47.8	1175	10	US-09-771-161A-224	Sequence 224, App
22	44	47.8	1175	10	US-09-771-161A-225	Sequence 225, App
23	44	47.8	1175	10	US-09-771-161A-226	Sequence 226, App
24	43	46.7	117	10	US-09-864-761-34970	Sequence 34970, A
25	43	46.7	390	10	US-09-925-302-625	Sequence 625, App
26	43	46.7	538	12	US-10-023-529-43	Sequence 43, Appli
27	43	46.7	538	12	US-10-023-523-43	Sequence 43, Appli
28	42	45.7	88	10	US-09-867-550-1482	Sequence 1482, Ap
29	42	45.7	135	10	US-09-949-192-21	Sequence 21, Appli
30	42	45.7	150	10	US-09-864-761-35638	Sequence 35638, A
31	42	45.7	324	10	US-09-746-801A-61	Sequence 61, Appl
32	42	45.7	345	10	US-09-919-497-61	Sequence 61, Appli
33	42	45.7	531	10	US-09-925-300-1444	Sequence 1444, Ap
34	42	45.7	2843	8	US-08-681-219-32	Sequence 32, Appli
35	41	44.6	139	10	US-09-864-761-45205	Sequence 45205, A
36	41	44.6	197	12	US-10-042-417-32	Sequence 32, Appli
37	41	44.6	365	10	US-09-876-187-4	Sequence 4, Appli
38	41	44.6	365	10	US-09-749-728B-15	Sequence 15, Appli
39	41	44.6	393	10	US-09-925-300-1578	Sequence 1578, Ap
40	41	44.6	461	9	US-09-764-868-765	Sequence 765, App
41	41	44.6	504	10	US-09-764-864-1173	Sequence 1173, Ap
42	41	44.6	595	10	US-09-738-973-187	Sequence 187, App
43	41	44.6	726	10	US-09-770-689A-4	Sequence 4, Appli
44	41	44.6	803	10	US-09-770-689A-2	Sequence 2, Appli
45	41	44.6	881	10	US-09-816-860A-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1  
US-09-924-654-4  
; Sequence 4, Application US/09924654  
; Patent No. US20020146712A1  
; GENERAL INFORMATION:  
; APPLICANT: Goold, Richard D.  
; APPLICANT: Akerblom, Ingrid E.  
; APPLICANT: Seillamer, Jeffrey J.  
; APPLICANT: Coleman, Roger  
; TITLE OF INVENTION: TUMOR SUPPRESSOR  
; FILE REFERENCE: PC-0049 CIP  
; CURRENT APPLICATION NUMBER: US/09/924,654  
; CURRENT FILING DATE: 2001-08-07  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: PERL Program  
; SEQ ID NO 4  
; LENGTH: 583  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. US20020146712A1 496071CD1  
US-09-924-654-4

Query Match	100.0%	Score 92;	DB 10;	Length 583;
Best Local Similarity	100.0%	Pred No. 4.8e-05;		
Matches	16;	Conservative 0;	Mismatches 0;	Indels 0;
Gaps	0;			
OY	1	PSHKSTQRPQPGRQ 16		
Db	550	PSHKSTQRPQPGRQ 565		
RESULT 2				
US-09-925-300-1079				
; Sequence 1079, Application US/09925300				
; Patent No. US20020151681A1				
; GENERAL INFORMATION:				
; APPLICANT: Craig Rosen,				
; APPLICANT: Steve Ruben				
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies				
; FILE REFERENCE: PA101				

```
; CURRENT APPLICATION NUMBER: US/09/925,300
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 1079
; LENGTH: 594
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (430)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-300-1079

Query Match      100.0%; Score 92; DB 10; Length 594;
Best Local Similarity 100.0%; Pred. No. 4.8e-05;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PSHKSTORPPPPQGRQ 16
   |||||
Db 561 PSHKSTORPPPPQGRQ 576

RESULT 3
US-09-737-149-33
; Sequence 33, Application US/09737149
; Patent No. US20020077466A1
; GENERAL INFORMATION:
; APPLICANT: Spaderna, Steven K
; APPLICANT: Quinn, Kerry E.
; APPLICANT: Shimkets, Richard A.
; APPLICANT: Muralidhara, Padigaru
; APPLICANT: Spytek, Kimberly A.
; TITLE OF INVENTION: Polypeptides and Nucleic Acids Encoding Same
; FILE REFERENCE: 15966-620 CIP
; CURRENT APPLICATION NUMBER: US/09/737,149
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 60/170,564
; PRIOR FILING DATE: 1999-12-14
; PRIOR APPLICATION NUMBER: 60/173,165
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: 60/173,544
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 60/174,404
; PRIOR FILING DATE: 2000-01-04
; PRIOR APPLICATION NUMBER: 60/174,962
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: 60/223,929
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 33
; LENGTH: 901
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-737-149-33

Query Match      53.3%; Score 49; DB 10; Length 901;
Best Local Similarity 66.7%; Pred. No. 27;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 5 STORPPPPQGRQ 16
   |||||
Db 401 ASARPPPPQGAQ 412

RESULT 4
US-09-737-149-33
; Sequence 33, Application US/09737149
; Patent No. US20020077466A1
; GENERAL INFORMATION:
; APPLICANT: Spaderna, Steven K
; APPLICANT: Quinn, Kerry E.
; APPLICANT: Shimkets, Richard A.
; APPLICANT: Muralidhara, Padigaru
; APPLICANT: Spytek, Kimberly A.
; TITLE OF INVENTION: Polypeptides and Nucleic Acids Encoding Same
; FILE REFERENCE: 15966-620 CIP
; CURRENT APPLICATION NUMBER: US/09/737,149
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 60/170,564
; PRIOR FILING DATE: 1999-12-14
; PRIOR APPLICATION NUMBER: 60/173,165
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: 60/173,362
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: 60/173,544
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 60/174,404
; PRIOR FILING DATE: 2000-01-04
; PRIOR APPLICATION NUMBER: 60/174,962
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: 60/223,929
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 33
; LENGTH: 901
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-737-149-33

Query Match      53.3%; Score 49; DB 10; Length 901;
Best Local Similarity 66.7%; Pred. No. 27;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 5 STORPPPPQGRQ 16
   |||||
Db 401 ASARPPPPQGAQ 412

US-10-106-534-2
; Sequence 2, Application US/10106534
; Patent No. US20020168668A1
; GENERAL INFORMATION:
; APPLICANT: Curtis, Rory A.J.
; APPLICANT: Millennium Pharmaceuticals, Inc.
; TITLE OF INVENTION: 14691, A Human Glutamate Receptor Family
; TITLE OF INVENTION: Member and Uses Therefor
; FILE REFERENCE: MPI01-042PIRM
; CURRENT APPLICATION NUMBER: US/10/106,534
; CURRENT FILING DATE: 2002-03-26
; PRIOR APPLICATION NUMBER: 60/279,086
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 998
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-106-534-2

Query Match      53.3%; Score 49; DB 9; Length 998;
Best Local Similarity 66.7%; Pred. No. 30;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 5 STORPPPPQGRQ 16
   |||||
Db 395 ASARPPPPQGAQ 406

RESULT 5
US-09-737-149-8
; Sequence 8, Application US/09737149
; Patent No. US20020077466A1
; GENERAL INFORMATION:
; APPLICANT: Spaderna, Steven K
; APPLICANT: Quinn, Kerry E.
; APPLICANT: Shimkets, Richard A.
; APPLICANT: Muralidhara, Padigaru
; APPLICANT: Spytek, Kimberly A.
; TITLE OF INVENTION: Polypeptides and Nucleic Acids Encoding Same
; FILE REFERENCE: 15966-620 CIP
; CURRENT APPLICATION NUMBER: US/09/737,149
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 60/170,564
; PRIOR FILING DATE: 1999-12-14
; PRIOR APPLICATION NUMBER: 60/173,165
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: 60/173,362
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: 60/173,544
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 60/174,404
; PRIOR FILING DATE: 2000-01-04
; PRIOR APPLICATION NUMBER: 60/174,962
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: 60/223,929
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 8
; LENGTH: 1043
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-737-149-8

Query Match      53.3%; Score 49; DB 10; Length 1043;
Best Local Similarity 66.7%; Pred. No. 32;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 5 STORPPPPQGRQ 16
   |||||
Db 401 ASARPPPPQGAQ 412
```

ORGANISM: Epstein-Barr virus  
SEQUENCE DESCRIPTION: SEQ ID NO: 4;  
US-10-036-729-4

Query Match 51.1%; Score 47; DB 9; Length 345;  
Best Local Similarity 53.3%; Pred. No. 19;  
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 2 SHKSTORPPPPQGRQ 16  
Db 248 SHQAAQAQPPPPPGTQ 262

RESULT 8  
US-09-864-761-41757  
; Sequence 41757, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aescmca-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/006666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/006667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/006664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/006669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/006665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/006668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/006663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/006662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/006661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/006670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 41757  
; LENGTH: 80  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC023490.4  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.8  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.1

Query Match 51.1%; Score 47; DB 9; Length 196;  
Best Local Similarity 63.6%; Pred. No. 10;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 SHKSTORPPPP 12  
Db 5 SHTTTTPPPPP 15

RESULT 7  
US-10-036-729-4  
; Sequence 4, Application US/10036729  
; Patent No. US20020169286A1  
; GENERAL INFORMATION:  
; APPLICANT: Middelidorp, Jaap Michiel.  
; TITLE OF INVENTION: Peptides and nucleic acid sequences  
; related to the Epstein-Barr virus.  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Akzo-No. US20020169286A1 Patent Department  
; STREET: 1300 Piccard Drive, Suite 206  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/036,729  
; FILING DATE: 21-Dec-2001  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/415,838  
; FILING DATE: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gortmley, Mary E.  
; REGISTRATION NUMBER: 34,409  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 345 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; ORIGINAL SOURCE:



ATTORNEY/AGENT INFORMATION:  
NAME: CERRONE, MICHAEL C.  
REGISTRATION NUMBER: 39,132  
REFERENCE/DOCKET NUMBER: PF-0535 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 855-0555  
TELEFAX: (650) 845-4166  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 274 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:

```

; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Flag Epitope
US-09-873-106B-22

Query Match 48.4%; Score 44.5; DB 10; Length 21;
Best Local Similarity 50.0%; Pred. No. 2.3;
Matches 9; Conservative 2; Mismatches 4; Indels 3; Gaps 1;

QY 1 PSHKSTQRPDPGGR 15
   | | | | | | | | | |
Db 4 PGRSQAPSHRPPPGHR 21

RESULT 14
US-09-873-106B-11
; Sequence 11, Application US/09873106B
; Patent No. US20020127657A1
; GENERAL INFORMATION:
; APPLICANT: Reinherz, Ellis L.
; APPLICANT: Freund, Christian
; APPLICANT: Li, Jing
; APPLICANT: Nishizawa, Kazuhisa
; APPLICANT: Wagner, Gerhard
; TITLE OF INVENTION: Cloning and Characterization of a CD2
; FILE OF INVENTION: Binding Protein (CD2BP2)
; FILE REFERENCE: 1062.1021-004
; CURRENT APPLICATION NUMBER: US/09/873,106B
; CURRENT FILING DATE: 2001-06-01
; PRIOR APPLICATION NUMBER: US 60/111,007
; PRIOR FILING DATE: 1998-12-04
; PRIOR APPLICATION NUMBER: US 60/115,647
; PRIOR FILING DATE: 1999-01-13
; PRIOR APPLICATION NUMBER: PCT/US99/26993
; PRIOR FILING DATE: 1999-11-15
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 70
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-873-106B-11

Query Match 48.4%; Score 44.5; DB 10; Length 70;
Best Local Similarity 50.0%; Pred. No. 7.7;
Matches 9; Conservative 2; Mismatches 4; Indels 3; Gaps 1;

QY 1 PSHKSTQRPDPGGR 15
   | | | | | | | | | |
Db 5 PGRSQAPSHRPPPGHR 22

RESULT 15
US-09-796-033-6
; Sequence 6, Application US/09796033
; Patent No. US20020009446A1
; GENERAL INFORMATION:
; APPLICANT: Magilav, Daniel
; TITLE OF INVENTION: METHOD OF MODULATING MEMORY EFFECTOR
; TITLE OF INVENTION: T-CELLS AND COMPOSITIONS
; FILE REFERENCE: 10274-044001
; CURRENT APPLICATION NUMBER: US/09/796,033
; CURRENT FILING DATE: 2001-02-27
; PRIOR APPLICATION NUMBER: PCT/US99/20026
; PRIOR FILING DATE: 1999-08-31
; PRIOR APPLICATION NUMBER: US 60/098,456
; PRIOR FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 351
; TYPE: PRT
; ORGANISM: Homo sapiens

```

; FEATURE:  
; NAME/KEY: SIGNAL  
; LOCATION: (1)....(24)  
US-09-796-033-6

Query Match 48.4%; Score 44.5; DB 10; Length 351;  
Best Local Similarity 50.0%; Pred. No. 40;  
Matches 9; Conservative 2; Mismatches 4; Indels 3; Gaps 1;

Qy 1 PSHKS---TORPPPPQGR 15  
| | | | : | | | |  
Db 286 PGRSQAPSHRPPPEGHR 303

Search completed: December 3, 2002, 14:29:56  
Job time : 7.46667 secs